Psychometrika

RELATIONS AMONG m SETS OF MEASURES*

PAUL HORST

UNIVERSITY OF WASHINGTON

The problem of determining linear functions for two sets of variables so as to maximize the correlation between the two functions has been solved by Hotelling. This article presents a more efficient computational solution for the case of two sets of variables and a generalized solution for any number of sets. Applications are discussed and a numerical example is included to demonstrate the solution for more than two sets.

I. *The Problem of Two Sets of Variables*

Suppose we have a set of n_1 predictor variables and a set of n_2 criterion variables for the same individuals. We wish to determine that linear combination of the predictor variables and that linear combination of the criterion variables which will yield the highest possible correlation between the two composites. Having determined these two linear functions, we wish to determine a second pair of linear functions which will yield two composites maximally correlated with each other but with the condition that each will correlate zero with each of the first pair of composites. We then seek a third pair of linear functions yielding maximally correlated composites but orthogonal to the first two pairs. This procedure may continue until we have n_1 or n_2 pairs, whichever is the smaller.

Another example where this type of analysis is useful is in factor analytic studies. Suppose a factor analysis has been conducted for the same set of variables on a group of normal individuals and on a group of mental hospital patients. It may be that the two factor matrices appear quite different but that transformations exist such that, for at least some of the factor vectors in one, closely corresponding factor vectors may be found in the other. As in the first example the restriction is imposed that the transformed vectors shall be orthogonal within each set, and, except for corresponding vectors, shall be orthogonal between sets.

The solution to the first problem was presented by Hotelling [3, 4].

The solution to the second problem is mathematically equivalent to that of the first and has been utilized by Wrigley and Neuhaus [11] and by Mees [5] following a development by Horst and Meredith ([5], pp. 101–109). Tucker [8] also considered a related problem in a method of inter-battery factor analysis.

A similar problem is encountered when a set of self-appraisal items is presented twice to a group with different instructions for each presentation. For example, the first time the subjects may be asked to respond as the items actually apply to them. The second time they may be asked to respond as they would wish the items to apply to them. Assume the items consist of subsets each purporting to assess a different personality dimension. By appropriate means a factor analysis can be conducted for the sets given under the two separate conditions. Assume the hypothesis that the subjects respond to the items relatively independently of the instructions. It should be possible, then, to find transformations such that the factor matrices for the two sets of instructions will be similar and satisfy the orthogonality conditions indicated in the previous examples. The solution for this problem is formally the same as for the other two examples.

A fourth example of the same type of problem is as follows. Suppose we have two sets of three tests each. For example the tests may be verbal, arithmetical, and spatial, but the tests measuring the same function in the two sets consist of different sets of items. Assume the two sets of tests are taken by the same group of persons. We wish to determine transformations for each set so that the transformed sets will all be mutually orthogonal except for corresponding variables between the two sets which should correlate as highly as possible.

In general, assume we have given the two sets of n_1 and n_2 measures, each for N entities. In the first example N is the number of cases, n_1 the number of predictors and n_2 the number of criteria. In the second example N is the number of variables, n_1 the number of factors for the normal group, and n_2 the number of factors for the hospital group. In the third example N is the number of variables and n_1 and n_2 are respectively the number of factors for the two conditions of administration. In the fourth example N is the number of cases and n_1 and n_2 are the number of tests in each set, namely three.

II. *The Solution for Two Sets of Variables*

The solution for the class of problems involving two sets of variables will now be outlined. Without loss of generality we may assume n_1 greater than or equal to n_2 . Let

$_1X$ be an $N \times n_1$ matrix of the first set of n_1 measures on the N entities,

$_2X$ be an $N \times n_2$ matrix of the second set of n_2 measures on the N entities,

$_1b$ be an $n_1 \times n_2$ transformation matrix to be determined,

$_2b$ be an $n_2 \times n_2$ transformation matrix to be determined.

Define

(1) $$_1Z = {}_1X \, _1b,$$

(2) $$_2Z = {}_2X \, _2b.$$

Let

(3) $$G_{11} = \frac{_1X'_1X}{N},$$

(4) $$G_{12} = \frac{_1X'_2X}{N},$$

(5) $$G_{22} = \frac{_2X'_2X}{N}.$$

In particular, if the X measures are all in standard units, the G matrices in (3), (4), and (5) are all correlation matrices. Also let

(6) $$\rho_{11} = \frac{_1Z'_1Z}{N},$$

(7) $$\rho_{12} = \frac{_1Z'_2Z}{N},$$

(8) $$\rho_{22} = \frac{_2Z'_2Z}{N}.$$

Using (1) through (8)

(9) $$\rho_{11} = {}_1b'G_{11} \, _1b,$$

(10) $$\rho_{12} = {}_1b'G_{12} \, _2b,$$

(11) $$\rho_{22} = {}_2b'G_{22} \, _2b.$$

According to the problem, the new Z variables should be uncorrelated except for corresponding variables in $_1Z$ and $_2Z$, which should be as highly correlated as possible. The variances of the Z variables are restricted, all equal to unity. Therefore ρ_{11} and ρ_{22} are identity matrices while ρ_{12} is a diagonal matrix whose diagonal elements are the correlations between the corresponding variables in $_1Z$ and $_2Z$.

The solution for $_1b$ and $_2b$ will be presented in a form equivalent to but computationally simpler than that given by Hotelling [3]. First let t_1 and t_2 be lower triangular matrices such that

(12) $$G_{11} = t_1 t_1' ,$$

(13) $$G_{22} = t_2 t_2' .$$

Define

(14) $$R_{12} = t_1^{-1} G_{12} t_2'^{-1} ,$$

and consider

(15) $$R_{12} = Q_1 \Delta Q_2' ,$$

where

(16) $$Q_1' Q_1 = Q_2' Q_2 = I ,$$

and Δ, without loss of generality, is diagonal with all diagonal elements non-negative and in descending order of magnitude from upper left to lower right.

Let

(17) $$M_1 = R_{12} R_{21} ,$$

(18) $$M_2 = R_{21} R_{12} ,$$

where R_{21} is the transpose of R_{12} . Then the latent roots of both M_1 and M_2 are given in Δ_1^2 ; Q_1 and Q_2 are matrices of the corresponding latent vectors of M_1 and M_2, respectively. The right side of (15) has been designated the basic structure of R_{12} ([1], ch. 18).

To get $_1b$ and $_2b$ first solve for Δ and Q_1 and Q_2 . Solutions for the latent roots and vectors of Gramian forms such as M_1 and M_2 are available, both for desk and electronic computers (e.g., Hotelling [2] and Wright [10], respectively). If n_2 is less than n_1 then Δ^2 and Q_2 would be solved from M_2 .

The solution for Q_1 from (15) is

(18a) $$Q_1 = R_{12} Q_2 \, \Delta^{-1} .$$

The solutions for the b matrices can then be shown to be

(19) $$_1b = t_1'^{-1} Q_1 ,$$

(20) $$_2b = t_2'^{-1} Q_2 .$$

It can be shown that ρ_{12} in (10) is precisely Δ. In the case of standard measures Δ_1 is the correlation between the first vectors of $_1Z$ and $_2Z$, Δ_2 the correlation between the next two, and so on. If one is interested in only the highest or the few highest Δ's it is not necessary to solve for the remaining ones. Iteration procedures [2, 10] for obtaining the Δ and Q matrices provide in order of magnitude the Δ's and their corresponding Q vectors, beginning with the largest Δ.

III. *The General Problem of m Sets of Variables*

The procedure as outlined is adequate when there are only two sets of variables. However, experimental situations may frequently be encountered where there are more than two sets. Suppose, in the second example, that instead of having the two factor loading matrices on the same tests for normal and hospital patients, there are separate factor matrices by sex for each group, or four factor matrices in all. These four matrices may appear quite dissimilar, depending on the methods of factoring used and the actual differences among the four groups. It is possible, however, that with suitable transformations of the four factor matrices they would all become more or less similar.

Suppose the personality items considered in the last section have been administered under more than two sets of instructions. For example, in addition to responding as previously indicated, the subjects may have been asked to respond in what they regarded as the socially desirable manner. There might then be a third set of factor loadings not necessarily similar to each of the first two. With suitable transformations, however, all three might become similar with respect to at least some of the factors.

Or again in the fourth example we might have not only two sets of similar test batteries administered to the same group of individuals but also a third battery, or even more. Again although the batteries did not yield similar results it is possible that transformations could be found so that at least some of the transformed variables were similar from one battery to another.

The general problem may now be stated. Suppose that for each of a set of N entities there are measures for m sets of attributes with n_i attributes in the ith set. Each entity may be a person and each attribute a test. Again each entity may be a test and each attribute may be a factor. In any case there would be m matrices of height N and width n_i . The matrices may be quite dissimilar. The n_i may vary from one matrix to another. We may, however, have reason to believe that transformations exist for the respective matrices such that the new matrices will be similar. The problem then is to find the m transformations which according to specified criteria will yield new matrices of maximum similarity. This problem is closely related to that of testing the independence of m sets of variates, considered by Wilks [9] and by Roy and Bargmann [6]. Since, however, the model outlined below imposes conditions which these authors have not included in their developments, their tests are not recommended in connection with this model. Further investigation may show that these tests or minor modifications of them would be appropriate. It is probable, however, that adequate tests would be somewhat more involved than those proposed by Wilks and by Roy and Bargmann.

IV. *A Solution for the General Case of m Sets of Variables*

A solution for the general case will now be indicated. Let

$_iX$ be the ith $N \times n_i$ matrix of measures for the ith set,

$_ib$ be the $n_i \times K$ corresponding transformation matrix where K will be taken as equal to the smallest n_i in the set of m,

$_iZ$ be the ith $N \times K$ transformed matrix.

Then

(21) $$_iZ = {}_iX \, _ib.$$

Let

(22) $$G_{ij} = \frac{_iX'_iX}{N},$$

(23) $$\rho_{ij} = \frac{_iZ'_iZ}{N},$$

(24) $$t_i t'_i = G_{ii},$$

(25) $$R_{ij} = t_i^{-1} G_{ij} t_i'^{-1}.$$

Because of (24) and (25)

(26) $$R_{ii} = I.$$

Let

(27) $$_iU = {}_iX t_i'^{-1}.$$

From (22), (25), and (27)

(28) $$R_{ij} = \frac{_iU'_iU}{N}.$$

Next, let

(29) $$_i\beta = t'_i \, _ib.$$

From (21), (27), and (29)

(30) $$_iZ = {}_iU_i\beta.$$

Define the following supermatrices:

(31) $$X = (_1X, \, _2X, \, \cdots, \, _mX),$$

(32) $$U = (_1U, \, _2U, \, \cdots, \, _mU),$$

(33) $$Z = (_1Z, \, _2Z, \, \cdots, \, _mZ),$$

$$(34) \qquad G = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1m} \\ G_{21} & G_{22} & \cdots & G_{2m} \\ \cdots\cdots\cdots\cdots\cdots \\ G_{m1} & G_{m2} & \cdots & G_{mm} \end{bmatrix},$$

$$(35) \qquad D_t = \begin{bmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 & \cdots & 0 \\ \cdots\cdots\cdots\cdots \\ 0 & 0 & \cdots & t_m \end{bmatrix},$$

$$(36) \qquad R = \begin{bmatrix} I & R_{12} & \cdots & R_{1m} \\ R_{21} & I & \cdots & R_{2m} \\ \cdots\cdots\cdots\cdots\cdots \\ R_{m1} & R_{m2} & \cdots & I \end{bmatrix},$$

$$(37) \qquad \rho = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2m} \\ \cdots\cdots\cdots\cdots\cdots \\ \rho_{m1} & \rho_{m2} & \cdots & \rho_{mm} \end{bmatrix},$$

$$(38) \qquad D_b = \begin{bmatrix} {}_1b & 0 & \cdots & 0 \\ 0 & {}_2b & \cdots & 0 \\ \cdots\cdots\cdots\cdots \\ 0 & 0 & \cdots & {}_mb \end{bmatrix},$$

$$(39) \qquad D_\beta = \begin{bmatrix} {}_1\beta & 0 & \cdots & 0 \\ 0 & {}_2\beta & \cdots & 0 \\ \cdots\cdots\cdots\cdots \\ 0 & 0 & \cdots & {}_m\beta \end{bmatrix}.$$

From (22), (31), and (34)

$$(40) \qquad G = \frac{X'X}{N}.$$

From (25), (27), (32), and (36)

$$(41) \qquad R = \frac{U'U}{N}.$$

From (23), (33), and (37)

$$(42) \qquad \rho = \frac{Z'Z}{N}.$$

From (25), (34), (35), and (36)

$$(43) \qquad R = D_t^{-1} G D_t'^{-1}.$$

From (29), (35), (38), and (39)

(44) $$D_t'^{-1} D_\beta = D_b .$$

From (21) through (25) and (40) through (44)

(45) $$\rho = D_\beta' R D_\beta .$$

We are now ready to determine the $_i\beta$ in (45) so as to satisfy certain criteria we may specify in ρ. First, observe that in the case of only two sets the solution considered in Section II determined $_1\beta$ and $_2\beta$ so that ρ_{11} and ρ_{22} in (37) are identity matrices, and so that ρ_{12} is diagonal with the diagonal elements in descending order of magnitude from upper left to lower right. This means that $_1\beta$ and $_2\beta$ are both orthonormal matrices. They are, in fact, Q_1 and Q_2 in (15).

At first blush we might think to seek a set of β's in (45) such that the diagonal submatrices of ρ will all be identity matrices and the off-diagonal submatrices would all be diagonal matrices. This would mean that we had achieved a set of Z matrices such that each Z matrix was orthonormal and all but corresponding vectors between any pair of sets were orthogonal. In further developments solutions will be restricted to the case where the diagonals of ρ are all unity. In the case of more than two sets, we cannot in general require that both the diagonal submatrices and the off-diagonal submatrices in ρ be diagonal.

First consider a solution for the first column vector of each of the $_i\beta$'s. Let $_i\beta_{.1}$ be the first column vector of $_i\beta$ and define

(46) $$D_{\beta_{.1}} = \begin{bmatrix} _1\beta_{.1} & 0 & \cdots & 0 \\ 0 & _2\beta_{.1} & \cdots & 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ 0 & 0 & \cdots & _m\beta_{.1} \end{bmatrix},$$

and

(47) $$_1\rho = D_{\beta_{.1}}' R D_{\beta_{.1}} .$$

The problem now is to determine the first columns of the $_iZ$ matrices so that they shall be as nearly similar as possible. We have already specified that the $_i\beta_{.1}$ be normal vectors so that the diagonal elements of $_1\rho$ in (47) are unity. The order of $_1\rho$ is of course m. If the X and hence the Z are in standard units then the off-diagonals of $_1\rho$ are correlation coefficients. In any case the more similar the corresponding $_iZ_{.1}$ vectors to one another, the more closely will the off-diagonals of $_1\rho$ approach unity. We may therefore specify that the sum of the elements in $_1\rho$ shall be a maximum with the condition that the $_i\beta_{.1}$ be normal. The function to be maximized therefore is

(48) $$\phi_1 = 1' {_1\rho} 1 - m,$$

where 1 is a unit vector. Let

(49) $$_1\lambda = {}_1\rho\, 1 - 1.$$

If we let $_1P = (R - I)$ the equation to be satisfied is (see Section VI)

(50) $$_1P\, D_{\beta.1}1 = D_{\beta.1}\, {}_1\lambda.$$

If $m = 2$ we have Hotelling's special case of two sets, and from (49) $\phi_1 = 2\, {}_1\rho_{12}$, i.e., the function maximized is simply twice the off-diagonal element of the second-order $_1\rho$ matrix or, in the case of standard measures in Z, simply the correlation between $_1Z_{.1}$ and $_2Z_{.1}$.

To solve for the $_i\beta_{.1}$ we use (50) iteratively. Begin with a first approximation to the $_i\beta_{.1}$, say $_1D_{\beta.1}$, and get

(51) $$_1P\, {}_1D_{\beta.1}1 = {}_1D_{B.1}1.$$

We let

(52) $$_1D_{1\lambda}^2 = {}_1D'_{B.1}\, {}_1D_{B.1},$$

(53) $$_2D_{\beta.1} = {}_1D_{B.1}\, {}_1D_{1\lambda}^{-1},$$

or in general

(54) $$_1P\, {}_kD_{\beta.1}1 = {}_kD_{B.1}1,$$

(55) $$_kD_{1\lambda}^2 = {}_kD'_{B.1}\, {}_kD_{B.1},$$

(56) $$_{(k+1)}D_{\beta.1} = {}_kD_{B.1}\, {}_kD_{1\lambda}^{-1}.$$

Computations (54) through (56) are repeated until $D_{\beta.1}$ and $D_{1\lambda}$ stabilize to any specified degree of decimal accuracy.

Next determine the second vectors of the $_i\beta$ matrices, that is, the $_i\beta_{.2}$. These will be chosen orthogonal to the corresponding $_i\beta_{.1}$ vectors. We let

(57) $$_2P = [I - D_{\beta.1}\, D'_{\beta.1}]\, {}_1P[I - D_{\beta.1}\, D'_{\beta.1}].$$

Then to solve for $D_{\beta.2}$, analogous to (50),

(58) $$_2P\, D_{\beta.2}1 = D_{\beta.2}\, {}_2\lambda,$$

and the iterative solutions analogous to (54), (55), and (56).

In general, the solution for the hth set of vectors $D_{\beta.k}$ out of the $_i\beta$ is

(59) $$_hP = [I - D_{\beta.(h-1)}\, D'_{\beta.(h-1)}]\, {}_{(h-1)}P[I - D_{\beta.(h-1)}\, D'_{\beta.(h-1)}],$$

(60) $$_hP\, D_{\beta.h}1 = D_{\beta.h}\, {}_h\lambda,$$

(61) $$_h\rho = D'_{\beta.h}\, {}_hP\, D_{\beta.h},$$

(62) $$D_{h\lambda}1 = ({}_h\rho - I)1 = {}_h\lambda,$$

with the kth iterative solution to satisfy (60) given by

$$(63) \qquad {_k}P_{(k-1)}D_{\beta.\lambda}1 = {_k}D_{B.\lambda}1,$$

$$(64) \qquad {_k}D_{\lambda\lambda}^2 = {_k}D'_{B.\lambda}\,{_k}D_{B.\lambda},$$

$$(65) \qquad {_{(k+1)}}D_{\beta.\lambda} = {_k}D_{B.\lambda}\,{_k}D_{\lambda\lambda}^{-1}.$$

TABLE 1

The Supermatrix G

	1	2	3	4	5	6	7	8	9
1	1.000	.249	.271	.636	.183	.185	.626	.369	.279
2	.249	1.000	.399	.138	.654	.262	.190	.527	.356
3	.271	.399	1.000	.180	.407	.613	.225	.471	.610
4	.636	.138	.180	1.000	.091	.147	.709	.254	.191
5	.183	.654	.407	.091	1.000	.296	.103	.541	.394
6	.185	.262	.613	.147	.296	1.000	.179	.437	.496
7	.626	.190	.225	.709	.103	.179	1.000	.291	.245
8	.369	.527	.471	.254	.541	.437	.291	1.000	.429
9	.279	.356	.610	.191	.394	.496	.245	.429	1.000

TABLE 2

The Diagonal Supermatrix D_t

	1	2	3	4	5	6	7	8	9
1	1.0000	0	0	0	0	0	0	0	0
2	.2490	.9685	0	0	0	0	0	0	0
3	.2710	.3423	.8997	0	0	0	0	0	0
4	0	0	0	1.0000	0	0	0	0	0
5	0	0	0	.0910	.9958	0	0	0	0
6	0	0	0	.1470	.2838	.9476	0	0	0
7	0	0	0	0	0	0	1.0000	0	0
8	0	0	0	0	0	0	.2910	.9567	0
9	0	0	0	0	0	0	.2450	.3739	.8945

V. *A Numerical Illustration of the Method*

To illustrate the method data are taken from Thurstone and Thurstone [7]. Table 1 is the G matrix, in this case a matrix of intercorrelations. The first three variables are tests designed to measure respectively, verbal, numerical, and spatial ability. The next two sets of three each are two other

TABLE 3

The Diagonal Supermatrix D_t^{-1}

	1	2	3	4	5	6	7	8	9
1	1.0000	0	0	0	0	0	0	0	0
2	-.2571	1.0325	0	0	0	0	0	0	0
3	-.2034	-.3928	1.1115	0	0	0	0	0	0
4	0	0	0	1.0000	0	0	0	0	0
5	0	0	0	-.0914	1.0042	0	0	0	0
6	0	0	0	-.1278	-.3008	1.0553	0	0	0
7	0	0	0	0	0	0	1.0000	0	0
8	0	0	0	0	0	0	-.3042	1.0452	0
9	0	0	0	0	0	0	-.1468	-.4369	1.1179

TABLE 4

The Supermatrix $D_t^{-1}G$

	1	2	3	4	5	6	7	8	9
1	1.0000	.2490	.2710	.6360	.1830	.1850	.6260	.3690	.2790
2	.0000	.9685	.3423	-.0210	.6282	.2230	.0352	.4493	.2958
3	.0000	.0000	.8997	.0165	.1583	.5408	.0481	.2415	.4814
4	.6360	.1380	.1800	1.0000	.0910	.1470	.7090	.2540	.1910
5	.1256	.6441	.3923	.0000	.9959	.2838	.0386	.5201	.3782
6	.0589	.0621	.5015	.0000	.0000	.9475	.0673	.2660	.3805
7	.6260	.1900	.2250	.7090	.1030	.1790	1.0000	.2910	.2450
8	.1952	.4930	.4238	.0498	.5341	.4023	.0000	.9567	.3739
9	.0588	.1398	.4431	-.0015	.1890	.3373	-.0001	.0000	.8945

sets of tests designed to measure the same functions. Thus the illustration used is the type considered in the fourth example in Section I but with three instead of only two sets of variables.

Table 2 gives the D_t supermatrix as defined by (24) and (35). Table 3

TABLE 5

The Supermatrix R, the Supervector $D_{\beta \cdot 1} 1$ and the Scalar Elements of $_1\lambda$

	1	2	3	4	5	6	7	8	9	$D_{\beta \cdot 1} 1$	Scalars of $_1\lambda$
1	1.000	.000	.000	.636	.126	.059	.626	.195	.059	.7323	1.4917
2	.000	1.000	.000	-.021	.633	.049	.035	.459	.129	.5139	
3	.000	.000	1.000	.016	.157	.521	.048	.238	.426	.4468	
4	.636	-.021	.016	1.000	.000	.000	.709	.050	-.002	.6586	1.4783
5	.126	.633	.157	.000	1.000	.000	.039	.532	.190	.6247	
6	.059	.049	.521	.000	.000	1.000	.067	.258	.299	.4195	
7	.626	.035	.048	.709	.039	.067	1.000	.000	.000	.6781	1.4993
8	.195	.459	.238	.050	.532	.258	.000	1.000	.000	.6395	
9	.059	.129	.426	-.002	.190	.299	.000	.000	1.000	.3621	

gives the supermatrix D_i^{-1}, while in Table 4 is the supermatrix $D_i^{-1} G$. Table 5 gives the R supermatrix defined by (43), together with the supervector $D_{\beta.1} 1$ and the $_1\lambda_i$ scalars.

Table 6 gives the supermatrix $_2P$ as defined by (59), together with vector

TABLE 6

The Supermatrix $_2P$, the Supervector $D_{\beta.2} 1$ and the Scalar Elements of $_2\lambda$

	1	2	3	4	5	6	7	8	9	$D_{\beta.2} 1$	Scalars of $_2\lambda$
1	0	0	0	.308	-.221	-.155	.250	-.171	-.167	-.6806	1.1075
2	0	0	0	-.274	.367	-.116	-.204	.226	-.016	.5743	
3	0	0	0	-.190	-.060	.387	-.175	.021	.292	.4550	
4	.308	-.274	-.190	0	0	0	.351	-.276	-.170	.7524	1.2371
5	-.221	.367	-.060	0	0	0	-.266	.254	.049	.5557	
6	-.155	-.116	.387	0	0	0	-.156	.055	.195	.3536	
7	.250	-.204	-.175	.351	-.266	-.156	0	0	0	-.7324	1.1392
8	-.171	.226	.021	-.276	.254	.055	0	0	0	.5477	
9	-.167	-.016	.292	-.170	.049	.195	0	0	0	.4045	

$D_{\beta_{.2}}$ 1 and the scalars $_2\lambda_i$. Table 7 gives the supermatrix $_3P$ as defined by (59) together with the supervector $D_{\beta_{.3}}$ 1 and the scalars $_3\lambda_i$. Table 8 gives the supermatrix D_β assembled from Tables 5, 6, and 7, and Table 9 gives the supermatrix ρ defined by (45).

TABLE 7

The Supermatrix $_3P$, the Supervector $D_{\beta_{.3}}$ 1 and the Scalar Elements of $_3\lambda$

	1	2	3	4	5	6	7	8	9	$D_{\beta_{.3}}$ 1	Scalars of $_3\lambda$
1	0	0	0	.000	.006	-.009	.000	.003	-.005	1 -.0228	.7318
2	0	0	0	.004	.162	-.248	-.010	.092	-.143	2 -.6372	
3	0	0	0	-.004	-.196	.299	.012	-.111	.173	3 .7703	
4	.000	.004	-.004	0	0	0	.000	.001	-.002	4 -.0122	.6298
5	.006	.162	-.196	0	0	0	-.005	.049	-.076	5 -.5485	
6	-.009	-.248	.299	0	0	0	.008	-.075	.116	6 .8361	
7	.000	-.010	.012	.000	-.005	.008	0	0	0	7 .0604	.4327
8	.003	.092	-.111	.001	.049	-.075	0	0	0	8 -.5395	
9	-.005	-.143	.173	-.002	-.076	.116	0	0	0	9 .8398	

Table 10 gives the supermatrix D_b as defined by (44). Table 11 is the same as Table 9 except that rows and columns are permuted so that the diagonal submatrices are those for which the sums of the intercorrelations have been maximized.

TABLE 8

The Diagonal Supermatrix D_β

	1	2	3	4	5	6	7	8	9
1	.7323	-.6806	-.0228	0	0	0	0	0	0
2	.5139	.5743	-.6372	0	0	0	0	0	0
3	.4468	.4550	.7703	0	0	0	0	0	0
4	0	0	0	.6586	-.7524	-.0122	0	0	0
5	0	0	0	.6247	.5557	-.5485	0	0	0
6	0	0	0	.4195	.3536	.8361	0	0	0
7	0	0	0	0	0	0	.6781	-.7324	.0604
8	0	0	0	0	0	0	.6395	.5477	-.5395
9	0	0	0	0	0	0	.3621	.4045	.8398

TABLE 9

The Supermatrix ρ

	1	2	3	4	5	6	7	8	9
1	1.000	.000	.000	.735	.030	-.021	.756	.022	.020
2	.000	1.000	.000	.024	.603	.002	-.025	.504	.039
3	.000	.000	1.000	-.016	-.037	.465	.016	.036	.267
4	.735	.024	-.016	1.000	.000	.000	.743	-.023	-.020
5	.030	.603	-.037	.000	1.000	.000	-.031	.635	-.039
6	-.021	.002	.465	.000	.000	1.000	.021	-.002	.165
7	.756	-.025	.016	.743	-.031	.021	1.000	.000	.000
8	.022	.504	.036	-.023	.635	-.002	.000	1.000	.000
9	.020	.039	.267	-.020	-.039	.165	.000	.000	1.000

TABLE 10

The Diagonal Supermatrix D_b

	1	2	3	4	5	6	7	8	9
1	.5093	-.9208	-.0157	0	0	0	0	0	0
2	.3551	.4142	-.9605	0	0	0	0	0	0
3	.4966	.5057	.8562	0	0	0	0	0	0
4	0	0	0	.5479	-.8484	-.0689	0	0	0
5	0	0	0	.5011	.4517	-.8023	0	0	0
6	0	0	0	.4427	.3732	.8823	0	0	0
7	0	0	0	0	0	0	.4304	-.9584	.1012
8	0	0	0	0	0	0	.5102	.3957	-.9308
9	0	0	0	0	0	0	.4048	.4522	.9388

TABLE 11

The Supermatrix $_{ij}\rho$

	1	4	7	2	5	8	3	6	9
1	1.000	.735	.756	.000	.030	.022	.000	-.021	.020
4	.735	1.000	.743	.024	.000	-.023	-.016	.000	-.020
7	.756	.743	1.000	-.025	-.031	.000	.016	.021	.000
2	.000	.024	-.025	1.000	.603	.504	.000	.002	.039
5	.030	.000	-.031	.603	1.000	.635	-.037	.000	-.039
8	.022	-.023	.000	.504	.635	1.000	.036	-.002	.000
3	.000	-.016	.016	.000	-.037	.036	1.000	.465	.267
6	-.021	.000	.021	.002	.000	-.002	.465	1.000	.165
9	.020	-.020	.000	.039	-.039	.000	.267	.165	1.000

VI. *Proof of the General Solution*

The proof of the method outlined in Section IV and illustrated in Section V is given below. First consider (48) and renumber it in sequence for convenient reference:

$$(66) \qquad \phi_1 = 1'\,_1\rho 1 - m.$$

Because of (47) and the definition of $_1P$ preceding (50), this may be written

(67)
$$\phi_1 = 1' \, D'_{\beta.1} \, {}_1P \, D_{\beta.1}1.$$

Impose the restriction

(68)
$$D'_{\beta.1} \, D_{\beta.1} = I,$$

and set up the function

(69)
$$\psi_1 = \phi_1 - 1' \, D'_{\beta.1} \, D_{\beta.1} \, {}_1\lambda$$

where $_1\lambda$ is a vector of Lagrangian multipliers. From (67) and (69) by symbolic differentiation with respect to $1'D'_{\beta.1}$,

(70)
$$\frac{\partial \psi}{\partial(1' \, D'_{\beta.1})} = {}_1P \, D_{\beta.1}1 - D_{\beta.1} \, {}_1\lambda,$$

or equating (70) to zero,

(71)
$${}_1P \, D_{\beta.1}1 = D_{\beta.1} \, {}_1\lambda,$$

which is the same as (50).

Premultiplying (71) by $D'_{\beta.1}$ and using (68)

(72)
$$D'_{\beta.1} \, {}_1P \, D_{\beta.1}1 = {}_1\lambda.$$

Substituting (47) in (72)

(73)
$${}_1\rho \, 1 - 1 = {}_1\lambda.$$

If the $Z_{.i}$ are in standard units then $_1\lambda_i$ is according to (73) simply the sum of the correlations involving $_iZ_{.1}$. From (66) and (73)

(74)
$$\phi_1 = 1' \, {}_1\lambda.$$

Therefore ϕ_1 is simply twice the sum of all the $_iZ_{.1}$ intercorrelations.

No rigorous proof has been developed to show that the iterative solutions given by (54), (55), and (56) converge for $D_{\beta.1}$ or that the ϕ_1 obtained is a maximum. However, suppose we let $D_{1\lambda}$ be a diagonal matrix such that

(75)
$$D_{1\lambda}1 = {}_1\lambda.$$

Substituting (75) in (71),

(76)
$${}_1P \, D_{\beta.1}1 = D_{\beta.1} \, D_{1\lambda}1.$$

Let

(77)
$$D_{1\lambda I} = \begin{bmatrix} {}_1\lambda_1 I & 0 & \cdots & 0 \\ 0 & {}_1\lambda_2 I & \cdots & 0 \\ \multicolumn{4}{c}{\dotfill} \\ 0 & 0 & \cdots & {}_1\lambda_m I \end{bmatrix},$$

where the identity submatrices in (77) have respectively the orders n_1, n_2, $\cdots$, n_m. Then it is obvious that

(78)
$$D_{\beta_{.1}} D_{1\lambda} = D_{1\lambda I} D_{\beta_{.1}}.$$

From (78) and (76)

(79)
$$_1P \, D_{\beta_{.1}} 1 = D_{1\lambda I} D_{\beta_{.1}} 1.$$

Now let

(80)
$$D_{1\lambda I} = dc,$$

where c is a scalar and d is a diagonal matrix such that

(81)
$$1' \, d \, 1 = 1.$$

From (79) and (80),

(82)
$$d^{-\frac{1}{2}} \, _1P \, d^{-\frac{1}{2}} \, d^{\frac{1}{2}} \, D_{\beta_{.1}} 1 = d^{\frac{1}{2}} \, D_{\beta_{.1}} 1c.$$

Suppose now

(83)
$$d^{\frac{1}{2}} \, D_{\beta_{.1}} = D_{F_{.1}}.$$

Because of (68), (80), and (81)

(84)
$$1' \, D'_{F_{.1}} \, D_{F_{.1}} \, 1 = 1,$$

(85)
$$d = \begin{bmatrix} (_1F'_{.1} \, _1F_{.1})I & 0 & \cdots & 0 \\ 0 & (_2F'_{.1} \, _2F_{.1})I & \cdots & 0 \\ \hline 0 & 0 & \cdots & (_mF'_{.1} \, _mF_{.1})I \end{bmatrix}.$$

Using (83) in (82),

(86)
$$[d^{-\frac{1}{2}} \, _1P \, d^{-\frac{1}{2}} - cI] \, D_{F_{.1}} \, 1 = 0.$$

From (86), d is determined so that a latent vector of $d^{-\frac{1}{2}} \, _1Pd^{-\frac{1}{2}}$ is precisely $D_{F_{.1}} \, 1$ and (85) is also satisfied. Also c is a latent root of this matrix. The iterative solution indicated by (54), (55), and (56) is equivalent to the Hotelling [2] method of solution with the additional feature that successive approximations are taken to the d matrix. It can readily be proved that the Hotelling iterative method converges to the largest latent root and the corresponding latent vector. It is interesting to note that because of (80) and (86), each submatrix R_{ij} in R is weighted relatively as the inverse geometric mean of its two corresponding $_1\lambda_i$ and $_1\lambda_j$ values. These $_1\lambda_i$'s, it is recalled, are the sums of the corresponding rows of $_1\rho$, minus unities.

Although the foregoing development does not constitute a rigorous proof of convergence it does appear to provide intuitive support. Furthermore

the experimental results so far obtained have converged, thus also providing empirical evidence.

Next we prove that (57) and (58) provide the solutions for the $_i\beta_{.2}$ vectors under the constraints that

(87) $$D'_{\beta_{.2}} D_{\beta_{.1}} = 0,$$

(88) $$D'_{\beta_{.2}} D_{\beta_{.2}} = I.$$

Let

(89) $$\phi_2 = 1' D'_{\beta_{.2}} {}_1P D_{\beta_{.2}} 1$$

and set up the function

(90) $$\psi_2 = \phi_2 - 1' D'_{\beta_{.2}} D_{\beta_{.1}} {}_{21}\gamma - 1' D'_{\beta_{.2}} D_{\beta_{.2}} {}_2\lambda,$$

where $_{21}\gamma$ and $_2\lambda$ are vectors of Lagrangian multipliers.

Differentiating (90) symbolically with respect to $1'D'_{\beta_{.2}}$ and equating to zero, because of (89),

(91) $$\frac{\partial \psi_2}{\partial(1' D'_{\beta_{.2}})} = {}_1P D_{\beta_{.2}} 1 - D_{\beta_{.1}} {}_{21}\gamma - D_{\beta_{.2}} {}_2\lambda = 0.$$

Premultiplying (91) by $D'_{\beta_{.1}}$ and using (87) and (88),

(92) $$D'_{\beta_{.1}} {}_1P D_{\beta_{.2}} 1 = {}_{21}\gamma.$$

Substituting (92) in (91),

(93) $$[I - D_{\beta_{.1}} D'_{\beta_{.1}}] {}_1P D_{\beta_{.2}} 1 = D_{\beta_{.2}} {}_2\lambda.$$

Because of (87), (93) can be written

(94) $$[I - D_{\beta_{.1}} D'_{\beta_{.1}}] {}_1P[I - D_{\beta_{.1}} D'_{\beta_{.1}}] D_{\beta_{.2}} 1 = D_{\beta_{.2}} {}_2\lambda,$$

or if

(95) $$_2P = {}_1[I - D_{\beta_{.1}} D'_{\beta_{.1}}] {}_1P[I - D_{\beta_{.1}} D'_{\beta_{.1}}],$$

then

(96) $$_2P D_{\beta_{.2}} 1 = D_{\beta_{.2}} {}_2\lambda,$$

which is the same as (58).

Finally we prove that (59) and (60) provide the solution for any $D_{\beta_{.k}}$ with the constraints that

(97) $$D'_{\beta_{.k}} D_{\beta_{.k}} = I$$

and

(98) $$D'_{\beta_{.k}} D_{\beta_{.k}} = 0$$

for every k less than h.

Let

(99)
$$\phi_h = 1' \, D'_{\beta.h} \, {}_1P \, D_{\beta.h} \, 1,$$

and set up the function

(100)
$$\psi_h = \phi_h - 1' \, D'_{\beta.h} \, D_{\beta.1} \, {}_{h1}\gamma - \cdots$$

$$- 1' \, D'_{\beta.h} \, D_{\beta.(h-1)} \, {}_{h(h-1)}\gamma - 1' \, D'_{\beta.h} \, D_{\beta.h} \, {}_h\lambda,$$

where the γ's and ${}_h\lambda$ are vectors of Lagrangian multipliers.

Differentiating (100) symbolically with respect to $1'D'_{\beta.h}$ and equating to zero,

(101)
$$\frac{\partial \psi_h}{\partial (1' \, D'_{\beta.h})} = {}_1P \, D_{\beta.h} \, 1 - D_{\beta.1} \, {}_{h1}\gamma - \cdots$$

$$- D_{\beta.(h-1)} \, {}_{h(h-1)}\gamma - D_{\beta.h} \, {}_h\lambda = 0.$$

Premultiplying (101) successively with the $D'_{\beta.h}$ for $k = 1, \cdots, (h-1)$ and using (97) and (98)

(102)
$$D'_{\beta.1} \, {}_1P \, D_{\beta.h} \, 1 = {}_{h1}\gamma,$$
$$\cdots$$
$$D'_{\beta.(h-1)} \, {}_1P \, D_{\beta.h} \, 1 = {}_{h(h-1)}\gamma.$$

Substituting (102) in (101)

(103) $$[I - D_{\beta.1} \, D'_{\beta.1} - \cdots - D_{\beta.(h-1)} \, D'_{\beta.(h-1)}] \, {}_1P \, D_{\beta.h} \, 1 = D_{\beta.h} \, {}_h\lambda.$$

Because of (98), (103) can be written as

(104)
$$[I - D_{\beta.1} \, D'_{\beta.1} - \cdots - D_{\beta.(h-1)} \, D'_{\beta.(h-1)}] \, {}_1P$$
$$\cdot [I - D_{\beta.1} \, D'_{\beta.1} - \cdots - D_{\beta.(h-1)} \, D'_{\beta.(h-1)}][D_{\beta.h} \, 1] = D_{\beta.h} \, {}_h\lambda.$$

But also because of (98),

(105)
$$[I - D_{\beta.1} \, D'_{\beta.1} - \cdots - D_{\beta.(h-1)} \, D'_{\beta.(h-1)}] = [I - D_{\beta.1} \, D'_{\beta.1}] \cdots$$
$$[I - D_{\beta.(h-1)} \, D'_{\beta.(h-1)}].$$

Therefore, if we define as in (59),

(106) $${}_hP = [I - D_{\beta.(h-1)} \, D'_{\beta.(h-1)}] \, {}_{(h-1)}P[I - D_{\beta.(h-1)} \, D'_{\beta.(h-1)}],$$

because of (105) and (106)

(107)
$${}_hP = [I - D_{\beta.1} \, D'_{\beta.1} - \cdots - D_{\beta.(h-1)} \, D'_{\beta.(h-1)}] \, {}_1P$$
$$\cdot [I - D_{\beta.1} \, D'_{\beta.1} - \cdots - D_{\beta.(h-1)} \, D'_{\beta.(h-1)}].$$

Using (107) in (104),

(108) $$_hP \ D_{\beta.h} \ 1 \ = \ D_{\beta.h} \ _h\lambda,$$

which is the same as (60).

REFERENCES

[1] Horst, P. Servant of the human sciences. Seattle: Division of Counseling and Testing Services, Univ. Washington, 1953. (Duplicated manuscript.)
[2] Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, 1933, **24**, 417–441, 498–520.
[3] Hotelling, H. The most predictable criterion. *J. educ. Psychol.*, 1935, **26**, 139–142.
[4] Hotelling H. Relations between two sets of variates. *Biometrika*, 1936, **28**, 321–277.
[5] Mees, H. L. Preliminary steps in the construction of factor scales for the MMPI. Unpublished doctoral dissertation, Univ. Washington, 1959.
[6] Roy, S. N. and Bargmann, R. E. Tests of multiple independence and the associated confidence bounds. *Ann. math. Statist.*, 1958, **29**, 491–502.
[7] Thurstone, L. L. and Thurstone, T. G. Factorial studies of intelligence. *Psychometric Monogr.*, No. 2, 1941.
[8] Tucker, L. R. An inter-battery method of factor analysis. *Psychometrika*, 1958, **23**, 111–136.
[9] Wilks, S. S. On the independence of *k* sets of normally distributed statistical variables. *Econometrica*, 1953, **3**, 309–326.
[10] Wright, C. E. Principal axis factor analysis program for the IBM Type 650. Tech. Rep., Office of Naval Res. Contr. Nonr-477(08), Univ. Washington Division of Counseling and Testing Service, 1957.
[11] Wrigley, C. and Neuhaus, J. O. The matching of two sets of factors. *Amer. Psychologist*, 1955, **10**, 418–419.
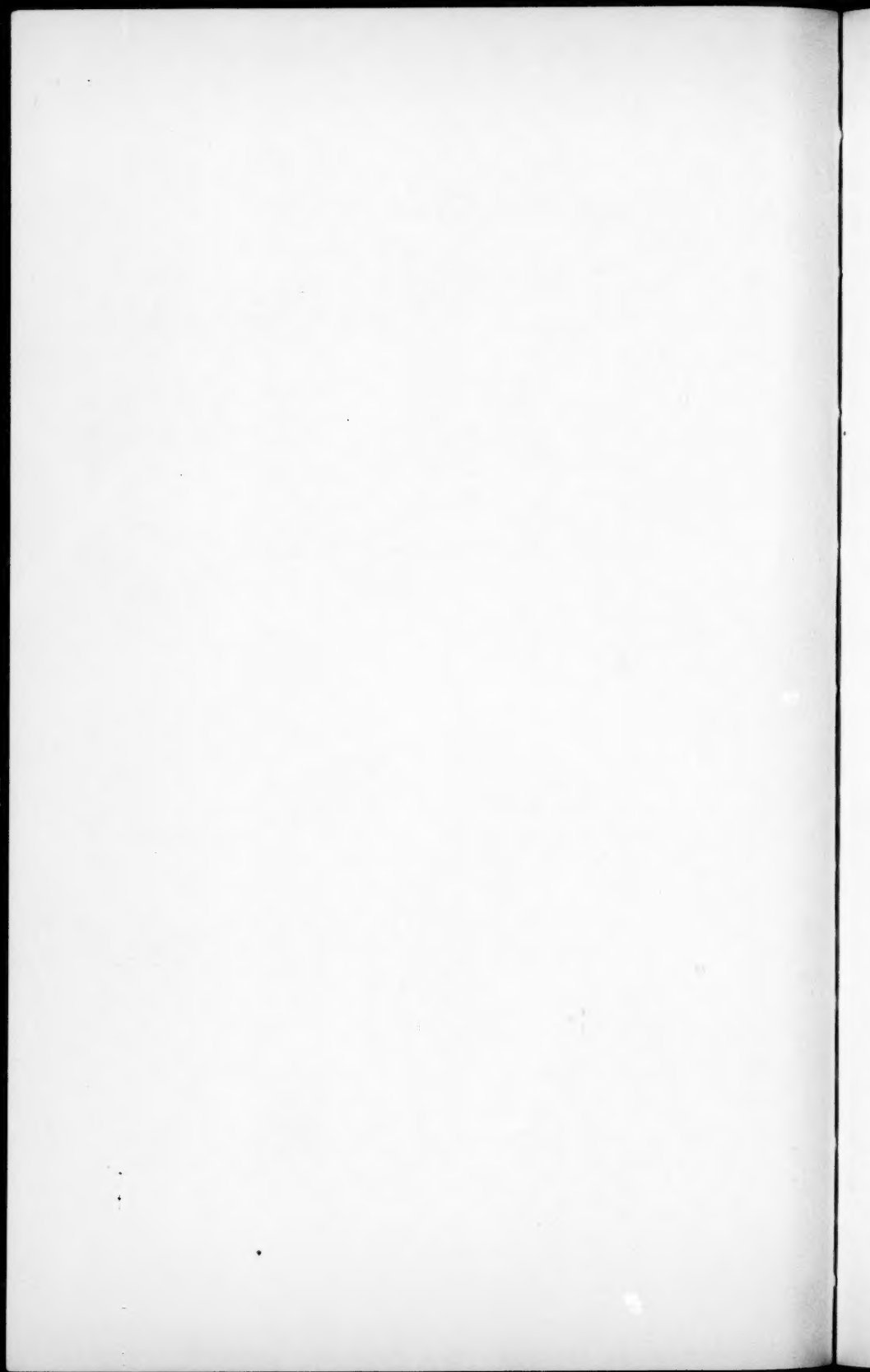
A CHOICE THEORY ANALYSIS
OF SIMILARITY JUDGMENTS*

R. DUNCAN LUCE

UNIVERSITY OF PENNSYLVANIA

The selection of one of several stimuli as most similar to a reference stimulus is assumed to satisfy a choice axiom that permits assigning ratio scale values to each variable-reference stimuli pair. The logarithm of this scale is treated as a distance measure, leading to the following testable conclusions about the pairwise choice probabilities as the reference stimulus is varied. First, the plot is a symmetrically truncated ogive with horizontal tails. Second, if two pairs of choice stimuli have the same midpoint, the ogive of one pair is part of the ogive of the other. In terms of this model, the hysteresis and midpoint displacement effects in the method of bisection are discussed, and relations with Coombs' unfolding techniques are explored.

The experimental technique to be considered is the trivial generalization of the complete method of triads [10] in which a subject is confronted by a finite set T of stimuli from which he must select one as "most similar" to a reference stimulus a. In the method of triads, T consists of only two stimuli.

For each subject in a given experiment and for every T and a, suppose a probability distribution $P_T(\cdot; a)$ governs his responses. Thus, $P_T(x; a)$ is the probability that, out of T, he selects x as most similar to a. With a held fixed and T treated as a variable, these are simply choice probabilities—not unlike those postulated in many models for ordinary discrimination experiments. Of the various theories that have been proposed to relate such choice probabilities one to another, the following choice axiom, which has been investigated in [5], is assumed.

If the probabilities are all different from 0 and 1, then for $x \in S \subset T$,

$$P_T(x; a) = P_S(x; a)P_T(S; a),$$

where

$$P_T(S; a) = \sum_{x \varepsilon S} P_T(x; a).$$

(The choice axiom can be stated to cover the case where some of the probabilities are 0 or 1, but we will confine ourselves to the more restricted case where they are different from 0 and 1.) An important, though simple, con-

sequence of this assumption is that a positive ratio scale exists over the alternatives which, via a simple formula, reproduces all of the probabilities. Because the stimulus a is a parameter in our problem, we must assume that the scale also depends upon a; hence, we write the scale value for stimulus x as $v(x, a)$. The theorem asserts that for $S \subset T$,

$$(1) \qquad P_S(x; a) = \frac{v(x, a)}{\sum_{y \in S} v(y, a)}.$$

In the remainder of this paper this assumption will be accepted as correct for similarity judgments, and several assumptions about $v(x,a)$ will be investigated. Actually, of course, our interest is in relations among the probabilities when a is varied. One hopes that such relations may be found because a stimulus can serve both as a reference and, in other presentations, as one of the comparison stimuli; however, it is not at all easy to guess the relations directly. Apparently it is simpler to assign a reasonable interpretation to v, then to impose assumptions upon v that seem plausible in the light of the interpretation, and finally to determine the restrictions thus implied on the probabilities themselves. Although this technique is familiar and has been used to advantage in the past, it is not at all evident to the writer exactly why it works.

Because the subject is asked to render a similarity judgment, it seems possible that $v(x, a)$ is some sort of measure of the similarity, or dissimilarity, of x and a. Of the two, it must be the first because, with a and $T-\{x\}$ fixed and x variable, $v(x, a)$ varies in the same direction as $P_T(x; a)$. Although there is no clear evidence or necessary reason, it is widely held that a measure of similarity must be in some sense symmetric. In this case, the immediate formalization that comes to mind is

$$(2) \qquad v(x, y) = v(y, x);^*$$

however, another possibility should also be considered. In contrast to most scales that have been studied in psychology, ours is a ratio scale, which for some purposes means that multiplicative inverses are appropriate symmetric pairs. It is not clear that this is wrong here, so one should also consider the assumption

$$(3) \qquad v(x, y) = 1/v(y, x).$$

We shall investigate both assumptions, first the latter and, after rejecting it, then the former.

---

*The following notational convention is employed. When a stimulus is to be considered fixed, letters such as a, b, $\cdots$ are used; when it is variable, x, y, $\cdots$ are used. Thus, (2) is not written $v(x, a) = v(a, x)$, as one might have expected, because we no longer want to consider a single, fixed reference stimulus.

The Assumption $v(x, y) = 1/v(y, x)$

Rewriting assumption (3) in the form $v(x, y)v(y, x) = 1$, then for four stimuli, w, x, y, z,

$$\frac{v(x, z)v(z, x)}{v(y, z)v(z, y)} = \frac{1}{1} = \frac{v(x, w)v(w, x)}{v(y, w)v(w, y)}.$$

Cross-multiplying,

$$\frac{v(x, z)v(z, x)}{v(y, z)v(w, x)} = \frac{v(x, w)v(z, y)}{v(y, w)v(w, y)}.$$

Let $P(x, y; z) = P_{\{x,y\}}(x; z)$, etc., then from (1) and the last equation,

$$(4) \qquad \frac{P(x, y; z)\, P(z, w; x)}{P(y, x; z)\, P(w, z; x)} = \frac{P(x, y; w)\, P(z, w; y)}{P(y, x; w)\, P(w, z; y)}.$$

The following "thought" experiment should convince the reader that (4) cannot be correct. Consider a unidimensional continuum such as sound intensity, and let x and y be stimuli chosen several jnds apart, with y more intense than x. Then choose z to be their midpoint in the sense that z is more intense than x but not as intense as y and $P(x, y; z) = \frac{1}{2}$. Finally choose w more intense than y and such that y is the midpoint of z and w, i.e., $P(z, w; y) = \frac{1}{2}$. Schematically, this situation is shown below.

$$\begin{array}{cccc} | & | & | & | \\ x & z & y & w \end{array}$$

Substituting these two values in (4),

$$\frac{P(z, w; x)}{P(w, z; x)} = \frac{P(x, y; w)}{P(y, x; w)}.$$

Because $P(w, z; x) = 1 - P(z, w; x)$ and $P(y, x; w) = 1 - P(x, y; w)$, it follows immediately that $P(z, w; x) = P(x, y; w)$. However, on the intensity continuum it was assumed that $x < z < w$, i.e., that z is closer than w to x, so one anticipates that $P(z, w; x) > \frac{1}{2}$. Equally well, $x < y < w$, so $P(x, y; w) < \frac{1}{2}$. But this contradicts what has just been shown to follow from (3). Although the experiment has not been done, the results seem so certain that one can safely reject the original assumption (3).

The Assumption $v(x, y) = v(y, x)$

A necessary and sufficient condition that (2) *hold over* x, y, z *is that*

$$(5) \qquad P(x, y; z)P(y, z; x)P(z, x; y) = P(x, z; y)P(z, y; x)P(y, x; z).$$

PROOF. Using (2),

$$1 = \frac{v(x, y)v(y, z)v(x, z)}{v(y, z)v(x, z)v(x, y)}$$

$$= \frac{v(x, y)v(y, z)v(z, x)}{v(z, y)v(x, z)v(y, x)}.$$

Substituting from (1) and cross-multiplying yields the result.

Conversely, $v(x, z)$ and $v(y, z)$ are determined by the choice probabilities up to an arbitrary multiplicative constant. Similarly, $v(x, y)$ and $v(z, y)$ are determined up to a multiplicative constant, which may be chosen so that $v(z, y) = v(y, z)$. Finally, the constant for $v(y, x)$ and $v(z, x)$ can be chosen so that $v(x, y) = v(y, x)$, leaving only the relation between $v(x, z)$ and $v(z, x)$ unspecified. But (5) ensures that they must be equal.

Although (5) is similar in form to an important condition implied by the choice axiom, derived in ([5], p. 16), they are actually logically independent.

No "thought" experiment seems to reject (5). This is not to say that the condition is correct, but only that if (2) is wrong, it is more subtly wrong than (3).

The notion of symmetry embodied in (2) will be assumed in the remainder of the paper.

Betweeness

The ideas in this section are closely related to nonprobabilistic notions in [3] and [6]. The continued use of the word "between" is justified because the present definition reduces to the usual one when the probabilities are 0 and 1. It should be noted that in terms of some definitions of distance, this definition does not preclude three stimuli forming certain types of triangles.

In the following definitions and results, probabilities of $\frac{1}{2}$ are excluded because such symmetric cases are difficult to handle neatly in stating results.

DEFINITION 1. Let x, y, and z be stimuli, then y is *between* x and z if and only if $P(y, x; z) > \frac{1}{2}$ and $P(y, z; x) > \frac{1}{2}$. Denote this as $x \mid y \mid z$.

DEFINITION 2. Three stimuli form a *similarity intransitivity*, or briefly, an *intransitivity*, if and only if for some labeling x, y, and z, $P(x, y; z) > \frac{1}{2}$, $P(y, z; x) > \frac{1}{2}$, and $P(z, x; y) > \frac{1}{2}$.

Given three stimuli, at most one is between the other two; and, if they do not form an intransitivity, and none of the pairwise probabilities is $\frac{1}{2}$, then exactly one is between the other two.

PROOF. Without loss of generality, suppose both that x is between y and z and that y is between x and z; then, by definition, $P(x, y; z) > \frac{1}{2}$ and $P(y, x; z) > \frac{1}{2}$. Adding, $1 < P(x, y; z) + P(y, x; z) = 1$, a contradiction.

Now, suppose that none of the probabilities is $\frac{1}{2}$ and that no stimulus is between the other two. With no loss of generality suppose $P(x, y; z) > \frac{1}{2}$. Because x is not between y and z, it follows that $P(z, x; y) > \frac{1}{2}$. And because z is not between x and y, $P(y, z; x) > \frac{1}{2}$. Thus, the three elements form an intransitivity, contrary to assumption, so one must be between the other two.

If three stimuli satisfy (5), then they do not form an intransitivity.

PROOF. Suppose, on the contrary, x, y, and z do form an intransitivity as in definition 2, then

$$\frac{P(x, z; y)}{P(z, x; y)} < 1, \qquad \frac{P(y, x; z)}{P(x, y; z)} < 1, \qquad \frac{P(z, y; x)}{P(y, z; x)} < 1,$$

so,

$$\frac{P(x, z; y)P(y, x; z)P(z, y; x)}{P(z, x; y)P(x, y; z)P(y, z; x)} < 1,$$

contrary to (5).

This last result establishes that the choice axiom plus the symmetry condition $v(x, y) = v(y, x)$ implies some degree of unidimensionality in the responses, at least in the sense that intransitivities of three stimuli are impossible. Actually, this observation is really little more than a precursor to stating the usual, much stronger notion of unidimensionality: given a distance measure, then for y between x and z the distance from x to z is the sum of the distances from x to y and from y to z. The crucial question, of course, is what is meant by distance. Again, two possibilities come to mind. First, because $v(x, y)$ becomes larger as x and y become more similar, v itself cannot be a measure of distance, but $1/v(x, y)$ could be. Second, because there is evidence from other sources that the logarithm of the v-scale acts much like the interval scales that arise in Fechnerian and Thurstonian scaling, and because these scales have, in one way or another, been treated as measures of distance, $-\log v$ is a possibility. It will be shown that the former interpretation is untenable; then the consequences of the latter for the unidimensional case will be examined.

The Assumption That $1/v$ Is a Distance Measure*

If $1/v$ is a distance measure, in the usual sense, then $1/v(x, x) = 0$, so

$$P(x, y; x) = \frac{1}{1 + [v(y, x)/v(x, x)]} = 1,$$

for any y, however similar to x. Although it is probably unnecessary to cite data to convince the reader that this is wrong, they do exist in [7].

The Assumption That $-\log v$ Is a Unidimensional Distance Measure

In order that $d(x, y) = -\log v(x, y)$ act like a measure of distance of a unidimensional continuum, it is necessary that

    (i) $d(x, y) = d(y, x)$,
    (ii) $d(x, y) \geq 0$ and $d(x, x) = 0$,
    (iii) if $x \mid y \mid z$, then $d(x, z) = d(x, y) + d(y, z)$.

*The following argument is due to Clyde Coombs; it is simpler than that originally used.

The first condition is guaranteed by the symmetry of v, equation (2). The second is satisfied if $v(x, x) = v(y, y)$, for all x and y, and $v(x, y) \leq v(x, x)$, for one may choose the unit of v so that $v(x, x) = 1$. The condition $v(x, x) = v(y, y)$ is equivalent to

$$(6) \qquad P(x, y; y) = P(y, x; x).$$

The third condition is equivalent to

$$(7) \qquad \text{if} \quad x \mid y \mid z, \quad \text{then} \quad v(x, z) = v(x, y)v(y, z).$$

It should be noted that if $v(x, z) = v(x, y)v(y, z)$ and if all three v's are <1, then $x \mid y \mid z$.

To investigate the probability implications of (6) and (7), focus attention upon $P(a, b; x)$, treating a and b as fixed stimuli and the reference stimulus as a variable. Assuming that the order between a and b is fixed, there are two cases depending upon whether x is between a and b or not. First, look at the case where x is outside the interval defined by a and b.

If (6) and (7) hold, then for $a \mid b \mid x$,

$$P(a, b; x) = P(a, b; b),$$

and for $y \mid a \mid b$ and $a \mid b \mid x$,

$$P(a, b; x) = 1 - P(a, b; y).$$

PROOF.

$$
\begin{aligned}
P(a, b; x) &= \frac{1}{1 + [v(b, x)/v(a, x)]} \\
&= \frac{1}{1 + [1/v(a, b)]} \\
&= \frac{1}{1 + [v(b, b)/v(a, b)]} \\
&= P(a, b; b),
\end{aligned}
$$

which proves the first statement.

To prove the second,

$$
\begin{aligned}
P(a, b; x) &= \frac{1}{1 + [v(b, b)/v(a, b)]} \\
&= \frac{1}{1 + [v(a, a)/v(b, a)]} \\
&= P(b, a; a) \\
&= 1 - P(a, b; a) \\
&= 1 - P(a, b; y).
\end{aligned}
$$

Next, consider the case where x is between a and b. To do so, a notion used earlier will be formalized.

DEFINITION 3.   The *midpoint* of stimuli a and b is that stimulus $\overline{ab}$ such that $a \mid \overline{ab} \mid b$ and $P(a, b; \overline{ab}) = \frac{1}{2}$.

It follows immediately that $v(a, \overline{ab}) = v(b, \overline{ab})$ and $P(a, \overline{ab}; b) = P(b, \overline{ab}; a)$.

Extending the betweeness notation in the obvious way, if $a \mid c \mid x \mid d \mid b$ and $\overline{ab} = \overline{cd}$, *then* $P(a, b; x) = P(c, d; x)$.

PROOF.   Because $a \mid c \mid \overline{cd} \mid d \mid b$, (7) and $v(c, \overline{cd}) = v(d, \overline{cd})$ imply

$$v(c, b) = v(c, \overline{cd})v(\overline{cd}, b)$$

$$= v(d, \overline{cd})v(\overline{ab}, b)$$

$$= v(d, \overline{ab})v(\overline{ab}, a)$$

$$= v(d, a)$$

$$= v(a, d).$$

But, by (7),

$$v(c, b) = v(c, x)v(x, b)$$

and

$$v(a, d) = v(a, x)v(x, d).$$

Hence,

$$\frac{v(c, x)}{v(d, x)} = \frac{v(a, x)}{v(b, x)},$$

from which the result follows by (1).

The empirical import of these two results is most easily seen graphically. If one considers pairs of stimuli having the same midpoint, then, independent of these stimuli, there is some function—presumably ogival—which determines $P(a, b; x)$ for $a \mid x \mid b$. For x outside this region, P is either the constant $\alpha = P(a, b; a)$ or $P(a, b; b) = 1 - \alpha$. See Fig. 1. Both aspects of this prediction should be possible to test experimentally.

If one only requires that $-\log v(x, y)$ act like a distance measure, not a unidimensional one, in the sense that part (iii) of the axiom is replaced by the triangle inequality, namely, for $x \mid y \mid z$, $d(x, z) \leq d(x, y) + d(y, z)$, then using much the same methods it is easy to show that

(i) $P(a, b; b) = 1 - P(a, b; a)$,

(ii)   for $a \mid b \mid x$, $P(a, b; b) \geq P(a, b; x)$,

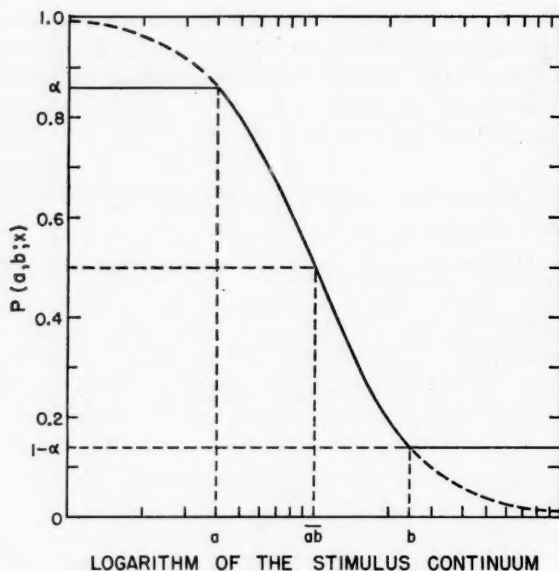(iii)   and for $y \mid a \mid b$, $P(a, b; a) \leq P(a, b; y)$.

FIGURE 1

Theoretical Plot of $P(a, b; x)$ as a Function of x

Thus, the effect of this change in assumptions is to round the corners of the function in Fig. 1 as the reference stimulus passes by a or b.

More detail about the form of $P(a, b; x)$ when x is between a and b can be determined by the following argument. When the reference stimulus x is extremely far from a and b, either above both or below both, the subject really only has to discriminate between a and b. For example, if beyond a doubt x is larger than both a and b, then he will report as most similar to x the one he believes to be larger, i.e., if $a \mid b \mid x$,

$$P(a, b) = P(a, b; x).$$

But it has just been shown that

$$P(a, b; x) = P(a, b; b).$$

So,

$$P(a, b) = P(a, b; b).$$

Thus, if one assumes the discrimination probabilities also satisfy the choice axiom and if one denotes the corresponding scale values by $v(x)$, then for $v(a) \leq v(b)$ and $a \mid b \mid x$

$$\frac{1}{1 + [1/v(a, b)]} = \frac{1}{1 + [v(b)/v(a)]}.$$

The other possible cases yield the same result, namely

(8)
$$v(a, b) = \begin{cases} v(a)/v(b) & \text{if } v(a) \leq v(b) \\ v(b)/v(a) & \text{if } v(a) \geq v(b). \end{cases}$$

Equation (8), then, establishes a basic connection between the discrimination and the similarity data if the present theory is correct. Indeed, similarity distance, $-\log v(a, b)$, is simply the absolute value of the difference of the logarithms of the discriminative scale values—what have been called Fechnerian scale values [5]. Thus, the model is substantially like Coombs' unfolding technique, where $-\log v(a, b)$ is the folded scale and $\log v(a)$ the unfolded one.

The existing evidence [7] is against the assumption $P(a, b) = P(a, b; b)$, but rather would suggest $P(a, b) < P(a, b; b)$. If one accepts the above argument for x sufficiently far from b and the discussion stemming from the triangular inequality, then

$$P(a, b) = P(a, b; x) < P(a, b; b),$$

and (8) is replaced by

$$v(a, b) > \begin{cases} v(a)/v(b) & \text{if } v(a) \leq v(b) \\ v(b)/v(a) & \text{if } v(a) \geq v(b). \end{cases}$$

Turning now to the case where $a \mid x \mid b$, then for $v(a) \leq v(x) \leq v(b)$, (8) implies

(9)
$$P(a, b; x) = \frac{1}{1 + [v(b, x)/v(a, x)]} = \frac{1}{1 + [v(x)^2/v(a)v(b)]}.$$

Bisection

In the psychophysical method of bisection the subject is required to adjust a variable stimulus until it is "half-way" between two other stimuli, a and b. It is plausible that he selects x so that a and b seem equally similar to it, in which case $x = \overline{ab}$, the midpoint of a and b. Thus, by (9)

$$P(a, b; \overline{ab}) = 1/2 = \frac{1}{1 + [v(\overline{ab})^2/v(a)v(b)]},$$

so

$$v(\overline{ab}) = [v(a)v(b)]^{1/2}.$$

That is to say, the discrimination v-scale value of the midpoint of two stimuli is the geometric mean of their discrimination v-scale values. One needs to convert this to a statement about the physical scale values.

Stevens [8] and Luce [4, 5] have argued that for at least certain classes of continua, the relation between a subjective scale, such as the v-scale, and the physical scale is a power function, i.e., $v(x) = \alpha x^\beta$. Thus, if

$$v(\overline{ab}) = [v(a)v(b)]^{1/2},$$

then

$$\alpha(\overline{ab})^\beta = [\alpha a^\beta \alpha b^\beta]^{1/2} = \alpha[(ab)^{1/2}]^\beta$$

so

$$\overline{ab} = (ab)^{1/2}.$$

Put in words, the physical scale value of the midpoint must also be the geometric mean of the physical scale values of the two stimuli, or, in a logarithmic transform of the physical scale—the corresponding db scale—it must be their average. It is well known that in general this is not correct [8, 9]. Not only is the subjective midpoint often shifted somewhat above the value just predicted, but its location differs depending upon whether the stimuli are presented in order a, x, b or b, x, a—this fact has been called a hysteresis effect.

Thus far any consideration of the well known fact that subjects exhibit response biases, often called time or space errors depending upon the mode of stimulus presentation, has been completely omitted. Possibly this can be used to explain the midpoint displacement and the hysteresis effects in the bisection method. Response biases will be treated in exactly the same way as in ([5], pp. 30–34).

Two distinct biases may enter. The first is due to the order of presentation of the stimuli; it affects their apparent intensities. Because the scale values can all be changed by a multiplicative constant without affecting (1), one of the biases may be chosen to be 1; let them be r, 1, and s for the first, second, and third presentations, respectively. Assuming that the ascending series is $a < x < b$ and the descending one, $b > y > a$, the intensity scale values are

Ascending: $v(a)r$, $v(x)$, $v(b)s$;
Descending: $v(b)r$, $v(y)$, $v(a)s$.

Thus, according to (8), the similarity scale values are

Ascending: $v(a, x) = v(a)r/v(x)$,
              $v(b, x) = v(x)/v(b)s$;
Descending: $v(a, y) = v(a)s/v(y)$,
              $v(b, y) = v(y)/v(b)r$.

The second bias arises if the subject has a differential tendency to set the middle stimulus nearer either the first or the last one presented. Let these biases be, respectively, 1 and t; so assume that x and y are chosen so

that

$$\text{Ascending: } v(a, x) = v(b, x)t,$$

which by previous equations is easily seen to be equivalent to

$$v(x) = v(\overline{ab})(rs/t)^{1/2};$$

and

$$\text{Descending: } v(a, y)t = v(b, y),$$

which is equivalent to

$$v(y) = v(\overline{ab})(rst)^{1/2}.$$

A hysteresis effect exists if and only if $v(x) \neq v(y)$, i.e., if and only if $t \neq 1$; it is of the sort observed, namely, $v(x) > v(y)$, if $t < 1$. Assuming $t = 1$, the bisection point differs from the midpoint provided $rs \neq 1$, and it is above the midpoint, as is generally observed, provided $rs > 1$. With $t \neq 1$, both bisection points are above the midpoint provided $rs > t$ and $> 1/t$. According to this model, the displacement from the midpoint and the hysteresis are independent biasing effects that one should be able to manipulate independently, e.g., by payoffs.

Strong Stochastic Transitivity

For a reference stimulus x, the condition of strong stochastic transitivity is:

if $P(a, b; x) > \frac{1}{2}$ and $P(b, c; x) > \frac{1}{2}$,

$$\text{then } P(a, c; x) \geq P(a, b; x), P(b, c; x).$$

Because the choice axiom has been assumed for a fixed reference stimulus, it is known ([5], p. 19) that this condition is satisfied.

Coombs [1, 2] in discussing preference data has argued that, at least in some cases, the choice between two stimuli really is determined by their similarity to some subjectively ideal stimulus on the continuum being judged, each subject having his own ideal. Thus, the present model for the method of triads, rather than the corresponding simple choice model, should apply to such data. Furthermore, Coombs has argued that if the subject fails to hold the ideal fixed, then apparent violations of strong stochastic transitivity can be expected to occur. This idea will now be examined in terms of the present model.

For the sake of simplicity, suppose that the variations of the ideal x are sufficiently small relative to the separations between the stimuli so that the order relations between the stimuli and the ideal are unchanged. As Coombs has pointed out, there are two inherently different cases. A uni-

lateral triple is a set of judged stimuli, $\{a, b, c\}$, which are all on one side of x, e.g., $a \mid b \mid c \mid x$. It is not difficult to show that variations in x, subject to the requirement that the order relations not change, cannot affect the strong stochastic transitivity property in this case.

Bilateral triples are of the form $a \mid x \mid b \mid c$. Suppose that $v(a) > v(x)$, $v(y), v(z) > v(b) > v(c)$. The first case considered here is what Coombs has called a bilateral adjacent triple:

$$P(a, b; x) > 1/2, \qquad P(b, c; y) > 1/2, \quad \text{and} \quad P(a, c; z) > 1/2.$$

Substituting

$$P(a, b; x) = \frac{1}{1 + [v(a)v(b)/v(x)^2]}, \quad \text{etc.,}$$

these three conditions are equivalent, respectively, to

$$v(x) > v(\overline{ab}), \qquad v(c) < v(b), \quad \text{and} \quad v(z) > v(\overline{ac}).$$

Violations of strong stochastic transitivity can occur in two ways:

$$P(a, c; z) < P(b, c; y), \text{ which is equivalent to } v(z) < v(\overline{ab}),$$

and

$$P(a, c; z) \qquad < P(a, b; x), \text{ which is equivalent to } v(z) < v(x) \left[\frac{v(c)}{v(b)} \right]^{1/2}.$$

The first violation appears to be easy to obtain provided the ideal is in the neighborhood of the midpoint $\overline{ab}$, for the only requirements are $v(x) > v(\overline{ab}) > v(z) > v(\overline{ac})$. The second appears much less likely to occur if c and b are not too close, for it requires a considerable shift in the ideals x and y. Coombs has found the first violation common, and the second much more rare in his data (personal communication).

The second case is that of a bilateral split triple

$$P(b, a; x) > 1/2, \qquad P(a, c; z) > 1/2, \quad \text{and} \quad P(b, c; y) > 1/2,$$

which are equivalent to $v(x) < v(\overline{ab})$, $v(z) > v(\overline{ac})$, and $v(c) < v(b)$. The possible violations are:

$$P(b, c; y) < P(b, a; x), \text{ which is equivalent to } v(x) < v(\overline{ac}),$$

and

$$P(b, c; y) < P(a, c; z), \text{ which is equivalent to } v(z) > v(\overline{ab}).$$

Thus, for the first to occur, the ideal must be located in the neighborhood of the $\overline{ac}$ midpoint and for the second it must be in the neighborhood of the $\overline{ab}$ midpoint. It is not obvious why this cannot happen, yet such violations are very rare in Coombs' data.

One may conclude, nonetheless, that the present theory is entirely consistent with Coombs' idea of violations in strong stochastic transitivity for certain types of data without, in fact, forcing one to reject the choice axiom.

REFERENCES

[1] Coombs, C. H. On the use of inconsistency of preferences in psychological measurement. *J. exp. Psychol.*, 1958, **55**, 1–7.

[2] Coombs, C. H. Inconsistency of preference as a measure of psychological distance. In C. W. Churchman and P. Ratoosh (Eds.) *Measurement: definitions and theories.* New York: Wiley, 1959. Pp. 221–232.

[3] Galanter, E. H. An axiomatic and experimental study of sensory order and measure. *Psychol. Rev.*, 1956, **63**, 16–28.

[4] Luce, R. D. On the possible psychophysical laws. *Psychol. Rev.*, 1959, **66**, 81–95.

[5] Luce R. D. *Individual choice behavior: a theoretical analysis.* New York: Wiley, 1959.

[6] Restle, F. A metric and an ordering on sets. *Psychometrika*, 1959, **24**, 207–220.

[7] Rosenblith, W. A. and Stevens, K. N. On the *DL* for frequency. *J. acoust. Soc. Amer.*, 1953, **25**, 980–985.

[8] Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

[9] Stevens, S. S. and Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

[10] Torgerson, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.
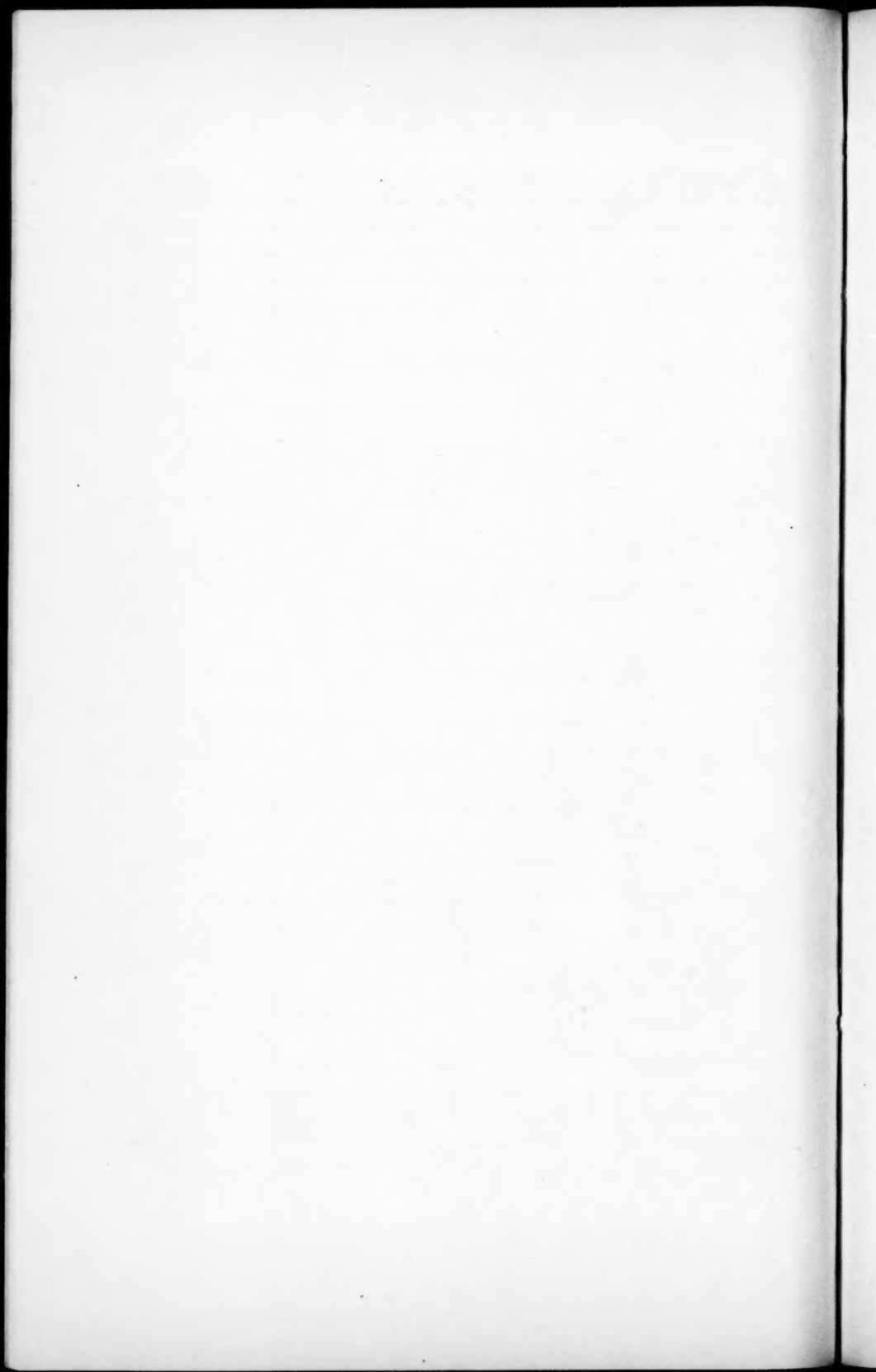
A DOUBLE LAW OF COMPARATIVE JUDGMENT FOR THE ANALYSIS OF PREFERENTIAL CHOICE AND SIMILARITIES DATA*

C. H. COOMBS, M. GREENBERG, AND J. ZINNES†

UNIVERSITY OF MICHIGAN

By virtue of certain modifications in the Law of Comparative Judgment, equations are developed which (i) permit the construction of a joint scale of individuals and items, as in the case of attitude measurement, directly from their pair-comparison preferences, and (ii) take into account the variable of laterality which is significant for the construction of group preference scales.

This paper is concerned with the theoretical implications that the unfolding model of preferential choice [1] has for the Law of Comparative Judgment [4]. By the unfolding model, each individual and stimulus is viewed in terms of a distribution of "discriminal processes" located in the same space. On each pair-comparison trial the individual is represented by a point from his distribution and each of the two stimuli by points from their respective distributions. The individual's preferential choice on a given trial is assumed to reflect which of the two stimulus points is nearer the individual's point on that trial.

That this theory has implications for the application of the Law of Comparative Judgment to pairwise preferential choice data has already been suggested by experimental results [3]. This experiment indicates that if the distributions of discriminal processes of the stimuli are both on the same side of, or unilateral to, the individual's distribution, the inconsistency of judgment is of a different order of magnitude than if the distributions are on opposite sides (bilateral). The consequence of this, in brief, is that the usual data matrix containing the proportion of times stimulus j has been preferred to stimulus k should be partitioned into two distinct data matrices. One matrix has proportions in each cell based on that subset of individuals for whom that pair of stimuli is unilateral, and the other matrix has proportions based on the subjects for whom that pair of stimuli is bilateral.

Because the relation of inconsistency to psychological distance is of a different order of magnitude for these two matrices, the Law of Comparative Judgment for each is different and hence gives rise to the reference in the title to a double Law of Comparative Judgment, a unilateral law for the unilateral matrix, and a bilateral law for the other. In the following section

the two laws are developed in order to show the theoretical implications of the unfolding theory for the Law of Comparative Judgment, and in the final section some of the implications for and difficulties of practical applications are pointed out.

Unilateral and Bilateral Equations

The unfolding theory of preferential choice postulates the existence of a space consisting of ideal points for individuals, denoted by c's, and points corresponding to stimuli, denoted by q's. Throughout the remainder of this paper it will be assumed that this space is one-dimensional, and it will be called a J or joint scale. The algebraic distance from the ideal point of individual i to the stimulus j, at the moment h, is then defined as

$$(1) \qquad p_{hij} = c_{hij} - q_{hij} \, .$$

In terms of this model, individual i will prefer stimulus j to stimulus k at the moment h if and only if

$$(2) \qquad |\, p_{hij} \,| - |\, p_{hik} \,| \leq 0.$$

Alternatively, the preferential choice of an individual at a given moment signifies which stimulus point is nearer his ideal point. Furthermore, the percentage of times that one observes stimulus j preferred to stimulus k, then, is the percentage of times that (2) obtains. From (1)

$$(3) \qquad |\, p_{hij} \,| - |\, p_{hik} \,| = |\, c_{hij} - q_{hij} \,| - |\, c_{hik} - q_{hik} \,|.$$

It is assumed in the following discussion that when an individual is judging a pair of stimuli at a given moment, only one ideal point is involved. Thus, for stimuli j and k

$$(4) \qquad c_{hij} = c_{hik} = c_{hi} \, .$$

The development of these equations may be pursued in the context of case I of the Law of Comparative Judgment (replications on a single individual) or case II (replications over individuals). Because the possible applications of these developments will more likely be in the context of case II, the assumption will be made in what follows that each of a number of individuals has responded once to every pair of stimuli. Adaptations to case I or to a combination of case I and II, in which each of a number of individuals responds a number of times to each pair, are relatively straightforward and add little to the theoretical implications from case II alone. Hence in what follows the subscript h will be dropped so that individually replicated judgments will not be explicitly considered.

It is necessary to distinguish between those pairs of stimuli which are unilateral with respect to a given individual and those pairs which are bilateral. A stimulus pair (j, k) is unilateral to the subject located at c, if *both*

stimuli have scale values, q_{ij} and q_{ik}, less than or greater than c_i. More simply, both stimuli lie on the same side of c_i. For a bilateral stimulus pair, the two stimuli are on opposite sides of c_i. Formally, stimuli j and k are unilateral to c_i if and only if

$$(5) \qquad (q_{ij} \lessgtr c_i) \Leftrightarrow (q_{ik} \lessgtr c_i).$$

Stimuli j and k are bilateral to c_i if and only if

$$(6) \qquad (q_{ij} \lessgtr c_i) \Leftrightarrow (q_{ik} \gtrless c_i).$$

The unilateral equations will be developed first. As is evident from (5), if $q_{ij} < c_i$ and $q_{ik} < c_i$ then the individual is to the right of both stimuli; this is called condition R. (If $q_{ij} > c_i$ and $q_{ik} > c_i$ then we have condition L.)

Consider first condition L. From (5), since c_i is less than q_{ij} and q_{ik}, (3) reduces to

$$(7) \qquad |\, p_{ij} \,|^L - |\, p_{ik} \,|^L = (q_{ij} - c_i^L) - (q_{ik} - c_i^L) = q_{ij} - q_{ik},$$

and similarly, for c_i to the right of this unilateral pair,

$$(8) \qquad |\, p_{ij} \,|^R - |\, p_{ik} \,|^R = q_{ik} - q_{ij}.$$

Equations (7) and (8) indicate that the preferential choice of an individual for one of two unilateral stimuli is mediated by the difference between the scale values of the stimuli on the joint scale. This immediately suggests that the preferential choices of those individuals unilateral to a pair of stimuli can be used to scale the stimuli on the joint scale.

To simplify matters, the well-known case V assumptions will be made, i.e., that the stimuli project normal distributions on the J scale with equal variances,

$$(9) \qquad q_{ij} \quad \text{is} \quad N(Q_j, \sigma_q^2),$$

and that the correlation, over individuals, between each pair of stimuli is a constant,

$$(10) \qquad r_{q_{ij}q_{ik}} = r_{qq} \quad \text{for all pairs} \quad j, k.$$

The unilateral Law of Comparative Judgment may then be written as follows:

$$(11) \qquad |\, P_j \,|^L - |\, P_k \,|^L = X_{kj}^L \sigma_q \sqrt{2(1 - r_{qq})} = Q_j - Q_k,$$

where X_{kj}^L denotes the normal deviate corresponding to the proportion of unilateral-left persons preferring stimulus j to k and

$$|\, P_j \,|^L = \mathop{E}_i |\, p_{ij} \,|^L.$$

There is, of course, an equivalent expression which may be written for the R condition but, as will be discussed in the next section, only one of these is necessary in application.

The development of the Law of Comparative Judgment equation for the bilateral case parallels the unilateral treatment. As is evident from (6), if $q_{ij} < c_i < q_{ik}$ or if $q_{ij} > c_i > q_{ik}$ then the individual is between the stimuli on the joint scale and the stimuli are bilateral to the individual, in which case (3) can be written as follows (one may assume $q_{ij} < c_i < q_{ik}$ without any loss of generality):

$$(12) \qquad |\, p_{ij}\,|^R - |\, p_{ik}\,|^L = (c_i^{R_j L_k} - q_{ij}) - (q_{ik} - c_i^{R_j L_k})$$

$$= 2c_i^{R_j L_k} - q_{ij} - q_{ik}\,,$$

where $c_i^{R_j L_k}$ denotes the c_i of those individuals who are to the right of stimulus j and to the left of stimulus k.

A comparison of (12) with (7) or (8) makes evident the source of the essential difference between unilateral and bilateral preference judgments. In the unilateral case preference is mediated by the difference between the two scale values of the stimuli, completely independent of the c_i's. In the bilateral case, on the other hand, the c_i's enter in a significant way, and in particular, it is evident that the variance of the differences $|\, p_{ij}\,|^R - |\, p_{ik}\,|^L$ includes among its components the variance of the $c_i^{R_j L_k}$.

To simplify a good deal of tedious algebra, one may make the same case V assumptions previously introduced into the unilateral case [see (9) and (10)] and in addition assume

$$r_{c_i q_{ij}} = r_{c_i q_{ik}} = r_{cq}\,.$$

The variance of the differences on the left-hand side of (12), called the bilateral comparatal variance, may be written

$$(13) \qquad \sigma_{jk}^2 = 4\sigma_c^2 - 8r_{cq}\sigma_c\sigma_q + 2\sigma_q^2(1 + r_{qq}),$$

where σ_c^2 is the variance of the c_i which are bilateral to the pair of stimuli j and k.

The bilateral Law of Comparative Judgment with case V assumptions may then be written as

$$(14) \qquad |\, P_j\,|^R - |\, P_k\,|^L = X_{kj}^B \sqrt{4\sigma_c^2 - 8r_{cq}\sigma_c\sigma_q + 2\sigma_q^2(1 + r_{qq})}.$$

This bilateral comparatal variance is distinctly different from the unilateral comparatal variance under the same assumptions since, from (11), it is evident that the unilateral comparatal variance, σ_u^2, is

$$(15) \qquad \sigma_u^2 = 2\sigma_q^2(1 - r_{qq}).$$

If σ_u^2 is set equal to *one* for the unit of measurement in the unilateral case the bilateral comparatal variance may have some value quite different from *one*. The bilateral pairwise percentages are generated not only on the basis of a different unit of measurement, but, as may be seen from (12), are

estimates of a different variable than unilateral pairwise percentages. Unilateral and bilateral pairwise preferential choices should therefore not be combined in the same probability matrix and analyzed by the Law of Comparative Judgment.

Applications

Equations (11) and (14), for unilateral and bilateral judgments, respectively, constitute what is here called the double Law of Comparative Judgment. It is clear from these equations that according to the unfolding model of preferential choice the inconsistency measure for unilateral and bilateral pairs of stimuli must be differently translated into psychological distance, and furthermore, the inconsistency measures represent different variables.

There are two practical consequences. One is the possibility that arises for constructing the joint scale (i.e., the C_i and Q_j values) directly from preferential choice data instead of the usual two-step procedure of scaling the stimuli first and then getting preferential choice data. The second consequence is a revised procedure for translating the pairwise probabilities from similarities data into measures of distance. Both of these practical consequences are discussed in order in more detail below.

Any application of this development requires an initial step: knowing the approximate order of the stimuli on the J scale. If this order is not known from a priori considerations it can be obtained by utilizing the unfolding technique, which would also provide the approximate locations of the subjects with respect to the stimuli. The most serious problems in locating individuals from inconsistent data tend to arise with individuals centrally located on the J scale, i.e., who have a maximum number of bilateral pairs of stimuli.

This is fortunate for constructing the J scale in that it is the unilateral Law of Comparative Judgment which is needed for that purpose and it uses data only from individuals unilateral to a pair of stimuli. The entries in the unilateral matrix are obtained by the following procedure. There are N^L individuals to the left of stimuli j and k of whom n_{jk}^L prefer j to k. Similarly there are N^R individuals to the right of stimuli j and k of whom n_{kj}^R prefer k to j. A combined estimate of the proportion of individuals who *judge k to be greater than j* is

$$(16) \qquad \frac{n_{jk}^L + n_{kj}^R}{N^L + N^R}.$$

If X_{kj} represents the normal deviate corresponding to the proportion in (16), it is clear from (11) that

$$(17) \qquad Q_k - Q_j = X_{kj}\sigma_q \sqrt{2(1 - r_{qq})}.$$

Thus case V of the unilateral Law of Comparative Judgment may be applied to scale the stimuli on the J scale, and each individual may be assigned to an interval corresponding to his preference ordering.

The second consequence of this development is concerned with the analysis of similarities data to scale the distances between pairs of stimuli. In the first place it must be evident that this model of preferential choice has certain characteristics in common with similarities data. An individual making a preferential choice is, according to this model, judging the relative similarity of the stimuli to a hypothetical ideal point, so his judgment reflects an order relation on a pair of distances. One sometimes scales these distances from preferential choice data in order to construct a scale of the stimuli from most to least preferred. One scales the distances from similarities data in order to apply a multidimensional psychophysical model [5]. The double Law of Comparative Judgment has implications for both of these kinds of data.

Just the construction of a preferability scale of the stimuli (i.e., the P_i values) based on the entire set of subjects will be discussed. This has generally been done by applying the Law of Comparative Judgment to the proportion of times the members of a group have preferred each stimulus to every other, without any distinction between individuals. However (14) calls this procedure into question because the bilateral comparatal variance of each pair of stimuli is a function, among other things, of the variance of those individuals' ideal points for whom that pair is a bilateral pair. This value will in general be different for every pair of stimuli. A solution involves some difficult estimation problems and/or strong assumptions. For example one might assume $r_{cq} = r_{qq} = 0$, letting $\sigma_c = m\sigma_q$, then (14) becomes

$$(18) \qquad P_j^R - P_k^L = X_{ki}^B \sigma_q \sqrt{2} \ \sqrt{2m^2 + 1}.$$

Letting $\sigma_q \sqrt{2} = 1$ for the unit of measurement in the unilateral case, a solution to (18) is possible for the matrix of bilateral data only if the parameter m is known. At present only crude methods are available for estimating it, and none is recommended.

The final solution then for the group scale of preferability would involve a weighted average of the solutions to (18), (11) based only on unilateral left cases, and the corresponding equation for unilateral right cases. This procedure is recommended only in the absence of any better alternative and serves primarily to indicate how very different the problem is from that assumed in the conventional procedure.

Another area of application of these methods that is most promising is the area of similarities data. The frequency with which an individual judges stimulus B or C to be most like A is formally equivalent to A's preferential choice for the nearer one. In this case, if a one-dimensional scale may be obtained by the unfolding technique [2], then the unilateral law is appli-

cable for constructing an interval scale. If one wishes to scale the distances between pairs of stimuli then both laws need to be applied, and in the bilateral case, under case V assumptions, $m = 1$, because the variance of the c values between two stimuli is itself the variance of a stimulus. In this case (18) becomes

$$(19) \qquad \qquad |\, P_i \,|^R - |\, P_k \,|^L = X^B_{ki} \sigma_q \sqrt{2} \ \sqrt{3}.$$

So if $\sigma_q \sqrt{2}$ is set equal to 1 for the unit of measurement for the unilateral matrix, whereas $\sigma_q \sqrt{2} \ \sqrt{3}$ is used for the same purpose for the bilateral matrix of proportions, then the X^B_{ki} values from the bilateral matrix must be multiplied by $\sqrt{3}$ before combining with the X_{ki} from the unilateral matrix to form a weighted average. This is because the unit of measurement used for the bilateral percentages is $\sqrt{3}$ times as large as the unit used for converting the unilateral percentages. This is only true when the stimuli lie on a one-dimensional scale. The generalization of the effect of laterality on the comparatal variance for stimuli in a multidimensional space, while simple in principle, presents estimation problems which have not yet been solved.

A theoretical analysis of pairwise preferential choices is made in the spirit of the Law of Comparative Judgment but from the point of view of the unfolding theory of preferential choice behavior. The analysis reveals that for every pair of stimuli, the subjects must be partitioned into those who are (i) to the left of both stimuli on the J scale, (ii) between them, and (iii) to the right of both. The comparatal variance is seen to be different for (ii) than for (i) and (iii). It is shown how partitioning of the Ss will permit construction of a J scale directly from the preferential choices but a group scale of preferability has no simple solution. The appropriateness of this development for similarities data as well as preferential choice is pointed out.

REFERENCES

[1] Coombs, C. H. *A theory of psychological scaling.* Engng Res. Bull. No. 34, Ann Arbor, Mich.: Univ. Michigan Press, 1952.
[2] Coombs, C. H. A method for the study of interstimulus similarity. *Psychometrika*, 1954, **19**, 183–194.
[3] Coombs, C. H. On the use of inconsistency of preferences in psychological measurement. *J. exp. Psychol.*, 1958, **55**, 1–7.
[4] Thurstone, L. L. *The measurement of values.* Chicago: Univ. Chicago Press, 1959.
[5] Torgerson, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.
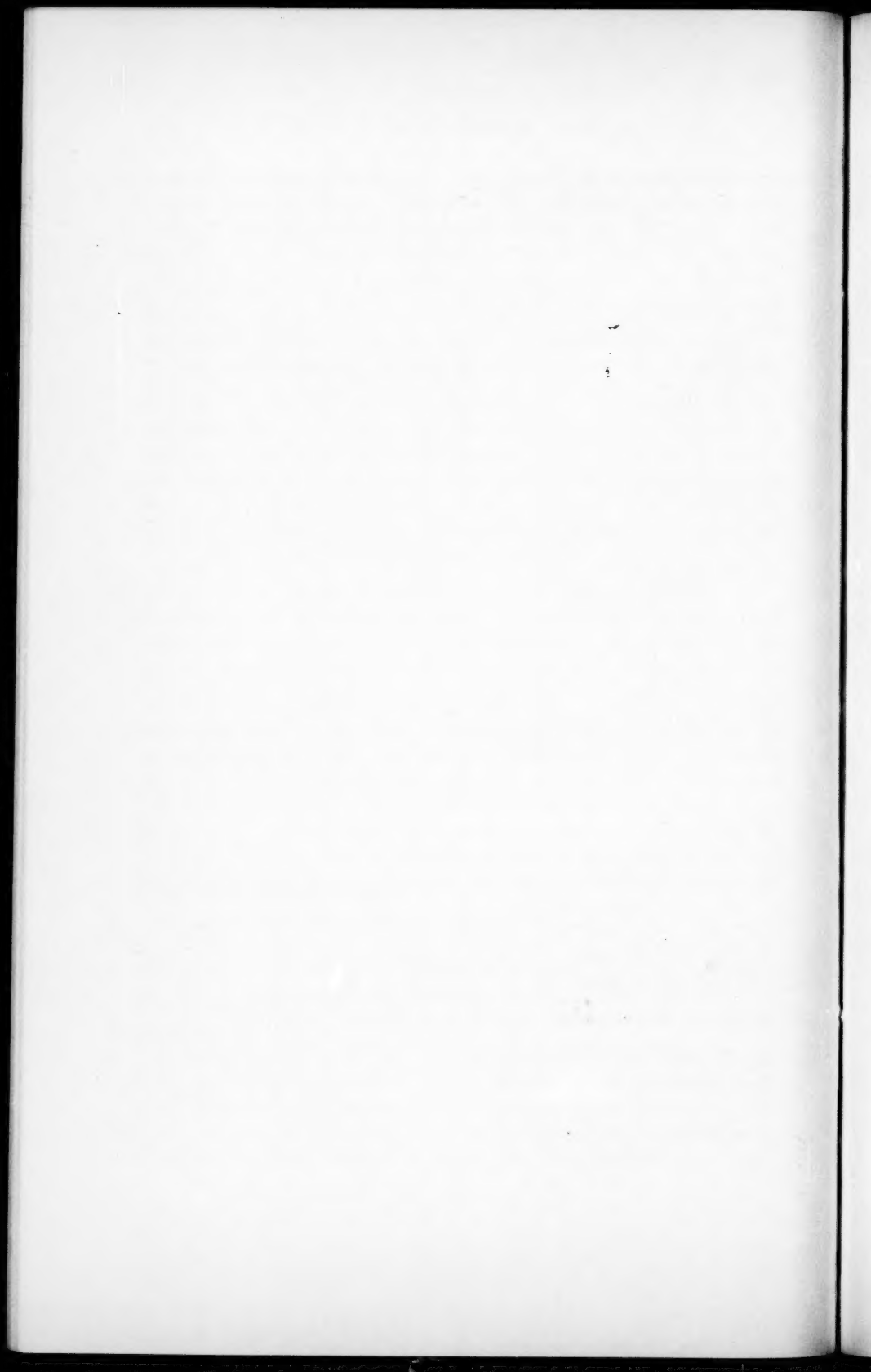
A GENERAL PROCEDURE FOR OBTAINING PAIRED COMPARISONS FROM MULTIPLE RANK ORDERS*

HAROLD GULLIKSEN AND LEDYARD R TUCKER†

PRINCETON UNIVERSITY AND EDUCATIONAL TESTING SERVICE

From a theoretical point of view, paired comparisons and the law of comparative judgment provide an excellent approach to the problem of psychological measurement. However, if a reasonably large number of stimuli are to be investigated, paired comparisons become extremely time-consuming and fatiguing to the subjects. A balanced incomplete block design, requiring multiple rank order judgments for each subject, provides an efficient experimental method for obtaining paired comparisons judgments. Features of the analysis proposed for this design are discussed in detail. A program for the analysis is available for the IBM 650 electronic computer.

The method of paired comparisons and the law of comparative judgment ([13], ch. 9) provide a scaling method with accurate checks on the goodness of fit of the data to the theory. Kendall ([9], ch. 11) has shown how the data for each subject may be analyzed to determine the number of departures from transitivity (i.e., the number of circular triads) in the subject's judgments. Transitivity in a paired comparison schedule means that for the pair AB we obtain the judgment $A > B$; for BC we obtain, say, $B > C$; and for AC we find $A > C$. As Kendall defines a circular triad, a departure from transitivity occurs since the subject judges $A > B$, $B > C$, and $C > A$. Thus, the method of paired comparisons is a very valuable one since it provides information on transitivity of preferences, on scale values, and on the applicability of a theory—the law of comparative judgment.

Use of the number of votes given to each stimulus by a subject as a type of ipsative score has interesting research possibilities. For example, the General Goals of Life Inventory prepared by Dunkel [6] for the Cooperative Study in General Education presented every pair of goals from a list of 20 goals and required paired comparison judgments by the subjects. The score on each goal for each subject was the number of times that subject preferred that goal to the other 19 goals. This instrument is in the general class of forced choice tests.

A computer program for complete paired comparisons data is described in [11] and i recorded in the IBM 650 Program Library—file no. 6.0.045.

From the subject's point of view, however, the method is laborious. For n stimuli, the number of pairs is $t = (n/2)(n - 1)$; for $n = 10$, 45 judgments are required; for 20 stimuli, $t = 190$; and for $n = 30$, $t = 435$. Studies using paired comparisons have usually been limited to 10 or 15 stimuli. Methods for using an incomplete paired comparison technique have been devised [10] and an experimental arrangement of the questionnaire that speeds up the judging process has been described by Benson [1]. Durbin [7] suggested the use of balanced incomplete block designs with rankings of objects within each block. He derived, after Kendall, a coefficient of concordance and a test of independence. However, Durbin did not carry his analysis through to a paired comparison scale. Independently, Bradley and Terry [2] have also suggested the use of block sizes larger than two. A method of analysis of variance and estimation of variance components for doubly balanced incomplete blocks using ratings instead of rankings has been presented by Calvin [3].*

The method of using multiple rank orders was devised to reduce the number of judgments required of the subject while obtaining information on each possible paired comparison. For twenty or thirty stimuli the rank order design takes, for each subject, only one-half to one-fourth the time required for the complete paired comparisons. Furthermore, the task of the subject is not extended in complexity up to ranking all objects in a single rank order. For 31 objects, the subject is asked to rank only 6 objects at any one time, 31 such rank orderings of length 6 being required. A single ranking of 31 objects is a formidable task for a subject [cf. 8]. The use of multiple rank orders permits observation of circular triads in the responses of each subject so that a count of the number of circular triads may be used as a measure of intransitivity for the subject.

These rank orders of six objects enable one to obtain all the paired comparisons, but sacrifice the possibility of obtaining information on circular triads, within each set of (say 6) objects. If all possible (15) pairs of the six objects were presented, the subject would then respond in one of the 2^{15} (or 32,768) possible ways. For the rank ordering of six objects there are only 6! (or 720) possible responses. For the complete paired comparison schedule of (say n) stimuli, the information is $(n/2)(n - 1)$. If the material is arranged in b blocks of k stimuli each, the amount of information is reduced to $b \log_2 k!$. Coombs [5] has utilized this method of evaluating the amount of information in various types of judgments.

An IBM 650 program has been written for processing data gathered by the multiple rank order method to facilitate the complete analysis of data. This program is recorded in the IBM 650 Program Library—file no. 6.0.038—and is presented in [12].

*The authors wish to thank an editorial reviewer for calling their attention to this work of Calvin's.

The method of using a balanced incomplete block design to give paired comparison information was suggested by Ledyard Tucker. The particular designs employed in the multiple rank order method were identified by John Tukey as balanced incomplete block designs of the sort described in Cochran and Cox [4]. A variety of such designs is available. The procedure for paired comparisons will be illustrated with a small set of seven stimuli, A–G. In the conventional paired comparisons method, the seven stimuli would be presented in 21 pairs. However, it is possible to present the seven stimuli in seven sets or items (designated a to g) of three each, as shown in Table 1. The subject is required to rank order each set of three, designating the best one by 1, and the poorest by 3. A possible set of rank orders for a hypothetical subject is also shown in Table 1.

Table 2 demonstrates how it is possible to obtain information on each

TABLE 1

Balanced Incomplete Block Design
For Seven Stimuli

a	b	c	d	e	f	g
A 3	A 1	G 2	D 3	E 2	C 2	F 2
B 1	D 3	A 3	B 1	G 3	D 1	C 3
C 2	E 2	F 1	F 2	B 1	G 3	E 1

TABLE 2

Information Given by B. I. B.
Design

	A	B	C.	D	E	F.	G
A	x	a	a	b	b	c	c
B	a'	x	a	d	e	d	e
C	a'	a'	x	f	g	g	f
D	b'	d'	f'	x	b	d	f
E	b'	e'	g'	b'	x	g	e
F	c'	d'	g'	d'	g'	x	c
G	c'	e'	f'	f'	e'	c'	x

of the 21 paired comparisons from the ranks given in Table 1. An x is placed in each diagonal cell in Table 2 since each object is not compared with itself, and, consequently, no data exist corresponding to these diagonal cells. The letter b is placed in the AD, AE, and DE cells of Table 2 because these pairs occur in item b. The letter b' is placed in the opposite cells DA, EA, and ED. Repeating this process for each of the seven items, we find that the arrangement in Table 1 will provide information on each of the 21 possible pairs.

Before administering any paired comparisons balanced incomplete block questionnaire, it is necessary to be certain that each pair of objects is presented together in a block or item exactly once. There must be no duplicate pairs and no omissions. The best way to be certain of this is to prepare, from the final copy of the questionnaire, a table such as is shown in Table 2, and to inspect it carefully to be certain that each cell has one entry. In this check only cells on one side of the diagonal need be tabulated and inspected.

Table 3 is prepared from the questionnaire, and shows the judgment

TABLE 3

Paired Comparisons Matrix

	A	B	C	D	E	F	G
A	x	1	1	0	0	1	1
B	0	x	0	0	0	0	0
C	0	1	x	1	1	1	0
D	1	1	0	x	1	1	0
E	1	1	0	0	x	0	0
F	0	1	0	0	1	x	0
G	0	1	1	1	1	1	x
V	2	6	2	2	4	4	1
R	7	3	7	7	5	5	8

for each pair. A one designates that the object at the top was ranked higher than the object at the side. A zero designates that the object at the side was ranked higher than the object at the top. For example, in item b, A is ranked higher than D, therefore, a zero is entered in the AD cell, and a one in the DA cell. The other 40 cells are filled in the same manner, an x being placed in the diagonal cells to indicate that no judgments were obtained in those cases. The row labeled V in Table 3 gives the sums of the columns, which is

the total number of "votes for" each of the stimuli or objects. For example, B received the maximum number of votes—six.

In row R of Table 3, we see how the information on the total number of votes for each of the objects may be readily obtained directly from the ranks. The row R shows the sum of the ranks assigned each of the objects in Table 1. The maximum possible sum of ranks occurs when the lowest rank (namely 3) is given to the object each of the three times it occurs, thus $9 = 3 \times 3$ is the largest number possible for the sum of the ranks. This number minus the sum of the ranks gives the votes for each of the objects. Kendall ([9], ch. 11) has shown how the number of circular triads, which he designates as d may be computed from the number of votes for each object. The formula he derives is

$$2d = (1/6)(n)(n - 1)(2n - 1) - \sum a_i^2 ,$$

where n is the number of objects compared (for the present illustration $n = 7$), and a_i is the votes for each object (a_i is shown in row V in Table 3). For the set of answers shown in Table 1, $d = (1/2)(91 - 81) = 5$, that is to say, there are five circular triads.

The section of the questionnaire with the multiple rank orders arranged in the balanced incomplete block design is termed the Rank-Order or (RO) section.

In preparing these questionnaires by a balanced incomplete block design, it has been found useful to provide for some assessment of an absolute standard for each subject. In order to do this, a complete list of the objects is presented at the end of the questionnaire, and the subject is asked to mark each item 1 if he likes it, and 2 if he dislikes it. The Like-Dislike may be changed into any two-choice category that is appropriate for the objects, and in harmony with the rank-order judgment, so that this "Like-Dislike" (or LD) section of the questionnaire may be included along with the rank order section and treated as if an $(n + 1)$th object (namely the zero point) has been added to the set. The number of objects checked "Dislike" would be interpreted as the number of "votes for the zero or neutral point." The "Like" votes are then distributed over the appropriate objects as an additional vote for that object, and the total circular triads count can be repeated.

By following a pattern analogous to that shown in Tables 1 to 3, any set of multiple rank orders presented in a balanced incomplete block design or in some other appropriate design (such as the balanced lattice) may be utilized to give paired comparison information. Appropriate designs are possible only with certain numbers of objects; however, "null objects" may be introduced and ranked lowest for the analysis, so that the available designs are generally usable for various numbers of stimuli.

Several designs that are useful for paired comparisons are presented by Cochran and Cox [4]. Certain balanced incomplete block designs can be

modified to give other suitable arrangements. Table 4 indicates some of the types of designs which seem to be the most useful for paired comparisons. It is possible to adapt any of these designs to *less* than the specified number of objects by simply introducing the required number of "null objects" and then assigning the lowest ranks to these objects in such a manner as to introduce no intransitivities from one block to another.

With regard to a statistical assessment of circular triads, Kendall has shown how to compute the maximum possible number of circular triads for paired comparisons. His formula is also correct for the balanced incomplete block design.

One method of evaluating the total circular triads d has been in terms of a coefficient of consistence ζ, defined

$$\zeta = 1 - \frac{d(\text{observed})}{d(\text{maximum})}.$$

As the number of circular triads varies from zero up to the maximum, the

TABLE 4

Some Rank Order Designs

Number of objects (n)	Number of blocks or items (b)	Number of objects in each item (k)	Code number for this design in reference [4]
7	7	3	11.7
9	12	3	10.1*
13	13	4	11.22
25	50	4	11.36
16	20	4	10.2*
21	21	5	11.34
25	30	5	10.3*
31	31	6	11.40
49	56	7	10.4*
57	57	8	11.44

* These are balanced lattice designs. A design of this type can be constructed from the design just below it by deleting one complete block of k items, thus reducing the number of blocks by one. Each of the objects, appearing in the deleted block, is then deleted from each of the remaining blocks, thus reducing the number of objects by k , and the number of objects per item by one. Thus, a computing program for a balanced incomplete block design can probably be used with only slight modification for the related balanced lattice design.

coefficient of consistence varies from unity down to zero. The maximum number of circular triads is given by

$$d(\max) = (n/24)(n^2 - 1), \qquad\qquad (n \quad \text{odd})$$

$$d(\max) = (n/24)(n^2 - 4), \qquad\qquad (n \quad \text{even}).$$

The work of Durbin [7] in determining the mean and variance of the coefficient of concordance for various incomplete block designs applies directly to evaluating total circular triads for multiple rank order designs in terms of a coefficient analogous to ζ. Durbin gives a formula for W, the coefficient of concordance,

$$(1) \qquad\qquad W = \frac{12S}{\lambda^2 n(n^2 - 1)} \, ,$$

where S is defined as the sum of the squared deviations from the mean of the total ranks assigned to each object i, that is,

$$(2) \qquad\qquad S = \sum r_i^2 - \frac{(\sum r_i)^2}{n} \, ,$$

where r_i is the sum of the ranks assigned to object i. For b blocks of k objects each

$$(3) \qquad\qquad \sum r_i = b(k/2)(k + 1).$$

Let us define a new coefficient

$$(4) \qquad\qquad Z = 1 - \frac{24 \, d}{n^3 - n}.$$

For n odd, Z is identical to the coefficient of consistence. For n even, Z varies from a maximum of unity (for d equal to zero) to a minimum of $3/(n^2 - 1)$. In this respect it is similar to Durbin's coefficient W as given in (1). For the special case in which $\lambda = 1$ and the rankings in each block are independent from those in other blocks, Durbin's coefficient of concordance W can be shown to be identical with Z for multiple rank order designs in which $\lambda = 1$.

To show that Z as defined by (4) is the same as W of (1), express S and d in common terms. The sum of the squares of the "votes for" scores for each object, designated $\sum a_i^2$ by Kendall, is related to the sum of the squared rank totals by

$$(5) \qquad\qquad \sum r_i^2 = \sum a_i^2 + (b^2 k^3/n).$$

Substituting (3) and (5) in (2) gives

$$(6) \qquad\qquad S = \sum a_i^2 - (b^2 k^2/4n)(k - 1)^2.$$

The suitable designs where $\lambda = 1$ are subject to the condition that the

number of object pairs $(n/2)(n - 1)$ is equal to the number of pairs in each block $(k/2)(k - 1)$ multiplied by the number of blocks b. Thus

(7) $$bk(k - 1) = n(n - 1).$$

Using (7) in (6)

(8) $$S = \sum a_i^2 - (n/4)(n - 1)^2.$$

Kendall has shown that

(9) $$2d = (n/6)(n - 1)(2n - 1) - \sum a_i^2 .$$

From (8) and (9) the relation between S and d is

(10) $$S = (n/12)(n - 1)(n + 1) - 2d.$$

Substituting (10) in (1) for the special case where $\lambda = 1$, and noting equation (4),

(11) $$W = 1 - \frac{24\, d}{n^3 - n} = Z.$$

It should be noted that, like W, Z varies from zero to unity for n odd and from $3/(n^2 - 1)$ to unity for n even.

Under the null hypothesis that the responses to each block are independently distributed from the responses to any other block, the expected values and variances for W given by Durbin also apply to Z for designs with $\lambda = 1$;

(12) $$E(Z) = \frac{k + 1}{n + 1} = 1 - \frac{n - k}{n + 1} ;$$

(13) $$\mathrm{Var}\,(Z) = \frac{2(k + 1)^2(n - k)}{n(n - 1)(n + 1)^2}.$$

When applied to the responses of a single individual to a multiple rank order questionnaire, this null hypothesis is tantamount to assuming that the individual is making his rankings at random.

From (4)

(14) $$d = \frac{n^3 - n}{24} - \frac{n^3 - n}{24} Z.$$

Using (12), (13), and (14),

(15) $$E(d) = (n/24)(n - 1)(n - k),$$

and

(16) $$\mathrm{Var}\,(d) = (n/288)(n - 1)(n - k)(k + 1)^2.$$

For the special case where $n = 31$ and $k = 6$,

$$E(d_{6,31}) = 968.75,$$

$$\text{Var } (d_{6,31}) = 3955.729167,$$

$$\text{St. Dev. } (d_{6,31}) = 62.894.$$

As indicated by Kendall and Durbin, the distribution of the coefficient of concordance W and hence of Z can be given by the beta distribution, or Fisher's F-ratio and (less accurately) by chi square. For random rankings the beta or Type I distribution gives

$$dF = \frac{1}{B(p, q)} W^{p-1}(1 - W)^{q-1} dW,$$

where

$$p = \frac{mn(1 - A)}{2\left(\dfrac{nm}{n - 1} - \dfrac{k}{k - 1}\right)} - A,$$

m denotes the number of times a given stimulus is ranked (thus $mn = bk$),

$$q = \left(\frac{1}{A} - 1\right)p,$$

where

$$A = \frac{(k + 1)}{\lambda(n + 1)}.$$

Fisher's variance ratio distribution may also be used, where

$$F = \frac{\left(\dfrac{\lambda(n + 1)}{k + 1} - 1\right)W}{1 - W},$$

when

$$\lambda = 1,$$

$$F = \frac{(n - k)W}{1 - W} = \frac{(n - k)Z}{(1 - Z)},$$

with degrees of freedom

$$\nu_1 = 2p \quad \text{and} \quad \nu_2 = 2q.$$

W also tends to be distributed as a multiple of chi square with $n - 1$ degrees of freedom

$$\chi^2 = \frac{\lambda(n^2 - 1)}{k + 1} W;$$

for $\lambda = 1$,

$$\chi^2 = \frac{n^2 - 1}{k + 1} Z.$$

In order to facilitate analyses of data collected by this particular method, an IBM 650 program has been prepared to handle the case in which 31 objects have been arranged in 31 blocks or items of 6 each, followed by the list of 31 objects for an absolute judgment of some type. This questionnaire design will be termed a "6-31 design." For *each subject* the program gives the following information.

1. The number of items checked liked, or positive, and the number disliked, or negative (from the LD section).
2. The number of votes for each of the 31 objects in comparison with the remaining 30 objects. Computationally this is given by subtracting the sum of the ranks from 36.
3. The number of circular triads in these votes for the 31 objects.
4. The number of votes for each of the 32 objects (e.g., the 31 specified objects plus an implied zero point, or neutral point, from the LD portion of the schedule).
5. The number of circular triads among the preferences for the 32 objects specified above.

For the *total group* of subjects, the program gives the following information.

1. For each paired comparison (i vs. j) the number and proportion of votes for i, and the number and proportion of votes for j. These have been designated respectively f' and p' by Torgerson ([13], p. 169 and p. 172).
2. The normal deviate (designated x' by Torgerson [13], p. 172), the arc sine and the logistic transform corresponding to each proportion.
3. The paired comparison scale values from complete data for each of the 31 objects. This is the value designated s'_i by Torgerson ([13], pp. 172–173). Three sets of scale values are given, one for the normal deviate transform, another for the arc sine transform, and a third for the logistic transform.
4. Three sets of paired comparison scale values from complete data for each of the "32 objects" obtained by considering the neutral point as another object. These, as in (3) preceding, are given in terms of the normal, the arc sine, and the logistic transforms.

This program will handle a maximum of 999 subjects in a single group. Each subject is processed in about 35 seconds; an additional 15 minutes is required for the computations involving the total group. Use of the program requires the minimum 650 installation, having a 2,000 word memory drum.

It is essential that the questionnaire data entered for this program be in proper form. That is to say, some permutation of the digits 1 to 6 must appear for each item in the rank order section, and a 1 or a 2 must appear for each answer in the Like-Dislike section. If there are omissions, or duplicate rankings, or inadmissible rankings, then peculiar, and possibly misleading results will probably be given by the program. In order to give a final check to the data cards before using them, an auxiliary checking program has been prepared, and is also included as a supplementary program with 6.0.038. Another program is being verified for handling 21 objects arranged in 21 items of 5 each—a 5–21 design—and will be made available through IBM when completed. Programs for the 4–13 and 4–25 designs are also being planned.

It is believed that the experimental convenience and efficiency of these balanced incomplete block designs with the accompanying programs, will make the balanced incomplete block design for paired comparisons a very generally useful experimental design for paired comparison studies.

REFERENCES

[1] Benson, P. Increasing the predictive efficiency of preference counts from paired comparison of personality traits. *Educ. psychol. Measmt*, 1958, **18**, 283–291.

[2] Bradley, R. A. and Terry, M. E. Rank analysis of incomplete block designs. The method of paired comparisons. *Biometrika*, 1952, **39**, 324–345.

[3] Calvin, L. D. Doubly balanced incomplete block designs for experiments in which the treatment effects are correlated. *Biometrics*, 1954, **10**, 61–88.

[4] Cochran, W. G. and Cox, G. M. *Experimental designs*. (2nd ed.) New York: Wiley, 1957.

[5] Coombs, C. A method for the study of interstimulus similarity. *Psychometrika*, 1954, **19**, 183–194.

[6] Dunkel, H. B. *General education in the humanities*. Washington: Amer. Counc. Educ., 1947.

[7] Durbin, J. Incomplete blocks in ranking experiments. *Brit. J. Psychol., Statist. Sect.*, 1951, **4**, 85–90.

[8] Hall, J. and Jones, D. C. Social grading of occupations. *Brit. J. Sociol.*, 1950, **1**, 31–55.

[9] Kendall, M. G. *Rank correlation methods*. (2nd ed.) London: Griffin, 1955. Ch. 11.

[10] Gulliksen, H. A least squares solution for paired comparisons with incomplete data. *Psychometrika*, 1956, **21**, 125–134.

[11] Gulliksen, H. An IBM 650 program for a complete paired comparisons schedule (Parcoplet 2–21). Tech. Rep. ONR contract Nonr 1858(15). (An *Appendix* to this technical report gives details on this program, including input-output formats and operating directions for this program which is recorded in the IBM 650 Program Library—file no. 6.0.045.)

[12] Gulliksen, H. and Tucker, L. R. An IBM 650 program for paired comparisons from balanced incomplete blocks—a 6–31 design (Parcobib 6–31). Res. Memo. 59-5. Princeton, N. J.: Educ. Testing Serv. ,1959. (This program is recorded in the IBM 650 Program Library—file no. 6.0.038.)

[13] Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.
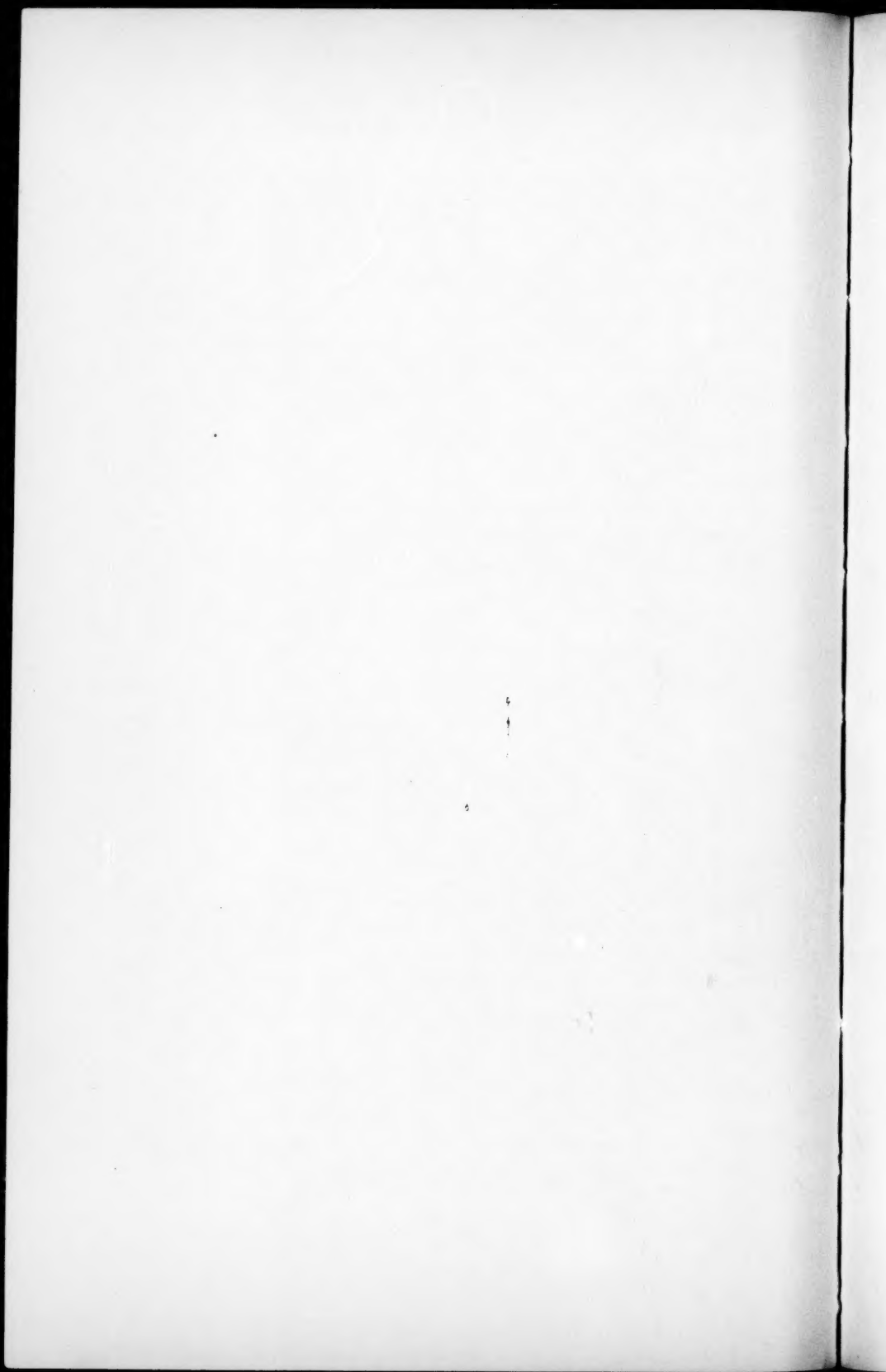
APPLICATION OF A TRACE MODEL TO THE RETENTION OF INFORMATION IN A RECOGNITION TASK

ROGER N. SHEPARD*

BELL TELEPHONE LABORATORIES

A stochastic model is proposed to account for the behavior of subjects in recognition tasks in which stimuli are presented, one at a time, in a protracted sequence. The basic assumption is that the memory trace resulting from the presentation of a particular stimulus not only fades away during the presentation of subsequent stimuli but also "diffuses" in such a way as to become decreasingly stimulus specific. An account is thereby provided for both (a) the increase in the probability of false recognition with the total number of stimulus presentations and (b) the departure of curves of forgetting from the previously proposed simple exponential decay functions. An expression for the amount of information carried along when the number of stimulus presentations becomes large is then derived for subjects who conform with the model.

The development of the model proposed here was primarily motivated by two considerations. First, attempts to obtain satisfactory estimates of the amount of information retained by subjects in certain recognition tasks without making any use of a substantive model for the memory process had been largely unsuccessful. Second, although a number of substantive models were already available for this purpose, they all seemed open to certain objections. In particular, most of these models predicted that forgetting always proceeds exponentially, whereas empirical curves often depart from a simple exponential decay function. Others failed to account for the fact that the probability that a new stimulus will be falsely recognized as old increases as more and more stimuli are presented. In the first part of this paper, then, a model for recognition memory is proposed that seems to overcome these difficulties. This model is essentially an extension of one originally developed by Shepard [18] to account for the shape of the gradient of stimulus generalization during paired-associate learning. In the second part of the paper, this model is then used as a basis for deriving the kind of informational estimate initially sought.

The type of experiment considered here is one in which stimuli are selected from a large population of, say, N stimuli and are presented to the subject one at a time. To simplify the exposition, it will be assumed that each stimulus is presented only once; i.e., the selection is without replacement.

(However, the extension of the model to cases in which stimuli may be presented any number of times is straightforward.) Immediately following the presentation of this sequence of stimuli, which will now be referred to as the *inspection sequence*, the subject is confronted with a *recognition array* consisting of some subset of R out of the N stimuli. The subject is instructed to indicate, for each of the R stimuli in the recognition array, whether it is "old" or "new" depending upon whether he does or does not recognize that stimulus as one already presented in the inspection sequence. The model relates the probability that a given stimulus in the recognition array will be classified as "old" to two independent variables: (*a*) the total number of stimuli in the inspection sequence and (*b*) the number of stimuli in that sequence intervening between the presentation of the given stimulus and the subsequent presentation of the recognition array.

Now the exponential curve of forgetting has typically been deduced by assuming (*a*) that the presentation of a stimulus leaves a trace in memory that is composed of a large number of separate elements and (*b*) that each of these elements has the same fixed probability of dropping out during any small fixed interval of time [3, 5, 13, 21, 24]. However, the empirical curves of forgetting often depart from this exponential in that they fall off too rapidly at first and then too slowly ([9], p. 25; [11], p. 609; [15], p. 22; [20], p. 350; [23], p. 373).

Another difficulty with theories that attempt to account for forgetting in terms of a simple attrition of elements is revealed by recognition experiments in which the stimuli vary along a single physical dimension. The procedure is to present one stimulus and then, after a certain delay, a second that differs from the first to some extent along the underlying physical dimension. The subjects are instructed to indicate whether the second stimulus is the same or different from the first [1, 2, 10, 12]. The finding is that the likelihood of the response "same" is greater when the second stimulus is separated from the first by a small rather than a large distance along the physical continuum. Moreover, as the delay increases, stimuli that are separated from the first by greater and greater distances along this continuum are likely to evoke the response "same." The trace seems not only to be fading away but also to be spreading out or diffusing along the underlying continuum. Certainly, then, to assume that the component elements simply drop out of the trace (as is sometimes proposed [13, 21]) is not sufficient. Some formulations (notably those of Estes [5] and Witte [24]) evidently do provide for a diffusion process. However, the implications of these formulations do not seem to have been worked out for recognition experiments involving large numbers of stimuli or stimuli that vary in similarity.

A theory is needed, then, to account for both the fading and spreading of the trace. Although such a theory could be confined to a specification of the functional relations between observable variables, the hypothetical trace elements will be retained in the present formulation. In this way the

apparent arbitrariness of the particular functions relating macroscopic variables can be removed by showing how these functions can arise as the combined effect of a large number of intuitively simpler micromechanical events. Assumptions about micromechanical processes can also serve as useful heuristics in exploring the relations between other macroscopic variables in new experimental contexts.

The Trace Model

Assumptions of the Basic Model

The stochastic model for stimulus generalization already proposed by Shepard [18] was based, essentially, on the following notions. To each stimulus, S_i, from some domain there corresponds a permanent internal representation, S_i^*. When S_i is presented, a set of activated trace elements is associated thereby with the representation S_i^*. Upon removal of the external stimulus, however, these trace elements do two things. (a) They gradually diffuse from S_i^* to other representations (such as S_k^*) corresponding to other stimuli that are similar to S_i. (b) They slowly become deactivitated (or in some other way leave the original domain). For purposes of recognition experiments, then, the probability that a particular stimulus would be recognized (i.e., classified as "old") is assumed to be a function of the number of trace elements associated with its internal representation. (The earlier formulation departs from this description in some minor respects, but it could be made compatible without altering its empirical consequences.)

No distinction will be made here between diffusion resulting solely from the passage of time and diffusion resulting from the presentation of subsequent stimuli. Since the stimuli are assumed to be presented at a relatively fixed rate, either interpretation can be made. The probability that a trace element associated with any representation S_i^* will transfer to some other representation S_k^* during a single presentation (or trial) of the inspection sequence will be denoted by v_{ik}. The micromechanical rules assumed to govern the redistribution of trace elements during each trial can then be set down as follows.

Assumption I. The diffusion of trace elements. For any pair of internal representations, S_i^* and S_k^*, there is a fixed probability v_{ik} that a trace element associated with one of these two representations will transfer to the other. This probability is, however, never greater than the probability that the trace element will remain associated with the same representation; i.e., $v_{ii} \geq v_{ik}$ for all i and k. (Diffusion is symmetric in the sense that $v_{ik} = v_{ki}$. Also, of course, $0 \leq v_{ik} \leq 1$ and $\sum_k v_{ik} = 1$.)

Assumption II. The deactivation of trace elements. Each activated trace element remains activated with probability u or becomes permanently deactivated (or, perhaps, leaves the domain) with probability $1 - u$. (Here, $0 \leq u \leq 1$.)

Assumption III. The introduction of new trace elements. A set of n newly activated trace elements is associated with the representation corresponding to the stimulus presented on that trial. (The exact value of n is unimportant; it is simply assumed to be sufficiently large that statistical fluctuations in the fraction of elements transferring from one representation to another can be disregarded.)

To specify the mechanics of the trace process is not alone sufficient; some assumption must also be made about the function relating the number of trace elements associated with an internal representation to the probability that the corresponding stimulus would be classified as "old." Under the condition that the subjects are not biased either toward responding "old" or responding "new," this function will simply be assumed to be one of direct proportionality. This is consistent with the assumption made in the earlier form of this model [18] (as well as with a similar assumption made in another connection by Luce [14], p. 23). However, owing to certain new features of the present application, special provisions are needed to set the factor of proportionality and to insure that the probability of response can not exceed unity. The explicit assumption, then, is as follows.

Assumption IV. The unbiased probability of response. In the absence of any response bias, the probability that a stimulus, S_i , presented in the recognition array will be classified as "old" is proportional to the number of trace elements associated with its internal representation, S_i^* . Moreover, the factor of proportionality is adjusted so that, for any given trial, the probability is maximum and equal to unity for the last stimulus in the inspection sequence; i.e., the subject is bound to recognize the stimulus immediately preceding the recognition array. (More generally, in order to prevent the probability from ever exceeding unity—as might otherwise happen if the same stimulus were presented more than once in the inspection sequence—the subject can be assumed to adjust the factor of proportionality so that the probability is unity for the stimulus corresponding to the internal representation with the largest number of trace elements.)

Now the present application of the model differs from the earlier one in that subjects are free to adjust their over-all tendency to classify stimuli as "old." Thus they might choose to classify a stimulus as "old" only if they are certain that they have seen that stimulus before; in which case they would also classify many old stimuli as "new." Or they might choose always to classify a stimulus as "old" unless they are certain that they have not seen that stimulus before; in which case they would also classify many new stimuli as "old." If the balance struck by subjects between these two types of errors is not under direct experimental control, an additional parameter would presumably be needed to accommodate possible response biases. (The term *bias* is used here in the sense of an asymmetric response tendency, not in the sense of an undesirable property of estimators.) Accordingly, a

final assumption must be made concerning the way in which the probability of response can be biased.

Assumption V. The response bias. The probability of actually classifying a stimulus as "old," denoted by P, is given in general by a simple power transformation of the unbiased probability, denoted by P^*; i.e.,

$$(1) \qquad P = (P^*)^B,$$

where $0 < B < \infty$.

The case in which the responses are not biased is now represented by setting $B = 1$ (the *null bias*). The choice of the power transformation here is admittedly arbitrary. However, it seemed to be the simplest function having the desirable properties (a) that any degree of bias can be achieved in either direction from the null bias and (b) that P always approaches 0 or 1 whenever P^* does so, regardless of the amount of bias. Thus, if a subject is certain that he has or has not seen a stimulus before, the bias will not prevent this information from determining his response. The bias transformation is illustrated, for representative values of the parameter B, in Fig. 1.
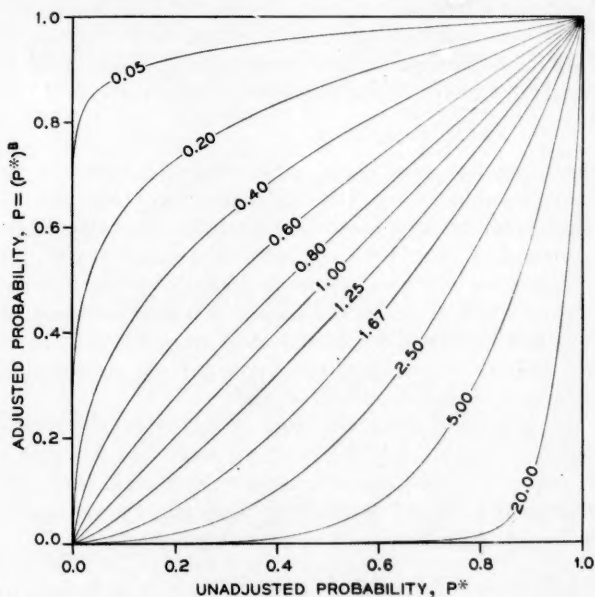

FIGURE 1

Illustrations of the Response-Bias Transformation for Representative Values of the Bias Parameter B

In addition to these five basic assumptions, two subsidiary provisos must also be stated. First, the number, R, of stimuli in the recognition array must be sufficiently small. If too many stimuli are included, the subject will have to scan them sequentially with the consequence that the traces remaining from the inspection sequence will dissipate still further and, hence, lead to spuriously low estimates of the probability of correct recognition. If necessary, this difficulty can be avoided by including only one stimulus in the recognition array. The entire set of N probabilities could still be estimated by confronting different subjects with different recognition arrays.

Second, the domain of stimuli cannot be chosen arbitrarily either. The model applies only to the extent that the population of stimuli from which the experimenter is actually drawing instances for presentation is coextensive with the effective set of stimuli from the standpoint of the subject. This means that the population of stimuli must be chosen to be psychologically circumscribed in the sense that the rate of diffusion of trace elements from representations inside to those outside the domain is negligible. For example, the set of three-digit numbers might constitute such a domain; for, if a subject has been shown only three-digit numbers, he probably will not make the error of classifying a four-digit number or a row of three letters as an "old" stimulus. In any case, a single number, N, will be used to represent both the number of stimuli in the experimenter's population and the number of internal representations that are significantly involved in the subject.

Diffusion of Traces in the Basic Model

Of primary interest, here, is the way in which the probability of recognition of a given stimulus decays while more and more other stimuli are subsequently presented. But information is also sought concerning the dependence of such a probability upon the total number of other stimuli presented in the entire inspection sequence. In order to examine the consequences of the trace model for these points of interest, it is first necessary to develop an equation showing how the distribution of trace elements depends upon the number of trials since a particular stimulus was presented, and upon the total length, T, of the inspection sequence.

Now the total number of activated trace elements associated with a representation S_i^* immediately following an inspection sequence consisting of T presentations (or trials) will be denoted by $n_i(T)$. The three rules governing the trace process (viz., I, II, and III) then imply a system of N simultaneous first-order linear difference equations; namely

$$(2) \qquad n_i(T) = u \sum_{k=1}^{N} v_k n_k (T-1) + n e_i(T), \qquad i = 1, 2, \cdots, N,$$

where

$$(3) \qquad e_i(T) = \begin{cases} 1 & \text{if } S_i \text{ was presented on trial } T \\ 0 & \text{otherwise.} \end{cases}$$

This system of equations has the matrix representation

$$(4) \qquad \mathbf{n}(T) = u\mathbf{V}\mathbf{n}(T - 1) + n\mathbf{e}(T),$$

where $\mathbf{n}(T)$ and $\mathbf{e}(T)$ are column vectors containing, respectively, the quantities $n_i(T)$ and $e_i(T)$ for $i = 1, 2, \cdots, N$, and where $\mathbf{V}$ is the symmetric matrix containing the $N \times N$ diffusion parameters v_{ik}. This equation can now be used to determine the effect of presenting a particular stimulus, say S_a, on the subsequent distribution of trace elements.

To this end it is convenient to contrast two possible experiments: one in which S_a is not presented at all; and one in which S_a is presented on some particular trial—say the dth from the last trial of the inspection sequence. Also, since it is only the effect of presenting S_a that is of concern here, the effects of all other presentations are best randomized. This can be accomplished, first, by considering that the T stimuli constituting the inspection sequence for each subject in both experiments are selected at random (except for the constraint on S_a itself) and, second, by contrasting the average distribution of activated trace elements resulting in one experiment with that resulting in the other. (In doing this the parameters v_{ik} and u are of course treated as identical for all subjects.)

First, the experiment is considered in which S_a is altogether excluded from the inspection sequence. On the assumption that the T stimuli are selected at random from the $N - 1$ stimuli excluding S_a, then, the difference equation for the expected change in the column vector $\mathbf{n}(T)$ after any one presentation becomes

$$(5) \qquad \mathbf{n}(T) = \mathbf{U}\mathbf{n}(T - 1) + \mathbf{c}(\bar{a}),$$

where $\mathbf{U} = u\mathbf{V}$, and where the elements $c_i(\bar{a})$ of the column vector $\mathbf{c}(\bar{a})$ are given by

$$(6) \qquad c_i(\bar{a}) = \begin{cases} 0 & \text{if } i = a \\ n/(N - 1) & \text{otherwise.} \end{cases}$$

Now the elements of the matrix $\mathbf{U}$ are positive and, since $u \sum_k v_{ik} = u < 1$, the sum of the elements in any row is less than 1. This implies that the matrix $(\mathbf{I} - \mathbf{U})$ is nonsingular and, hence, possesses an inverse ([8], p. 238). Therefore (5) has the solution

$$(7) \qquad \mathbf{n}(T) = \mathbf{U}^T\{\mathbf{n}(0) - (\mathbf{I} - \mathbf{U})^{-1}\mathbf{c}(\bar{a})\} + (\mathbf{I} - \mathbf{U})^{-1}\mathbf{c}(\bar{a}),$$

as can be verified by an induction on T.

Next the experiment is considered in which S_a is presented on the dth from the last trial of the inspection sequence. In this case d denotes the *delay* (in terms of the number of intervening presentations) between the presentation of S_a and the subsequent insertion of the recognition array. Since the trace elements introduced on different trials dissipate independently,

the solution for this case differs from that for the case in which S_a is not presented at all only in the activated residue of those elements that were introduced on the dth from the last trial. The column vector giving the distribution (after all T presentations) of just this residue will be denoted by $\mathbf{n}_d$. Then,

$$(8) \qquad \mathbf{n}_d = \begin{cases} \mathbf{U}^d \mathbf{c}(\bar{a}) & \text{if} \quad S_a \quad \text{is not presented on that trial} \\ \mathbf{U}^d \mathbf{c}(a) & \text{if} \quad S_a \quad \text{is presented on that trial,} \end{cases}$$

where the elements of the column vector $\mathbf{c}(a)$ are given by

$$(9) \qquad c_i(a) = \begin{cases} n & \text{if} \quad i = a \\ 0 & \text{otherwise.} \end{cases}$$

Thus the solution for the case in which S_a is presented on the dth from the last trial differs from the solution for the case in which S_a is not presented at all (7) only by the addition of the term $\mathbf{U}^d[\mathbf{c}(a) - \mathbf{c}(\bar{a})]$.

In applying the trace model to actual experiments it may often be reasonable to assume that the subjects have not been exposed to stimuli from the experimental domain for a long period preceding the experiment. In such cases the initial distribution of trace elements can be disregarded and $\mathbf{n}(0)$ dropped from (7). The general solution for the case in which the presentation of S_a and the subsequent presentation of the recognition array are separated by d other stimuli can then be put in the form

$$(10) \qquad \mathbf{n}(T) = (\mathbf{I} - \mathbf{U}^T)(\mathbf{I} - \mathbf{U})^{-1}\mathbf{c}(\bar{a}) + \mathbf{U}^d[\mathbf{c}(a) - \mathbf{c}(\bar{a})].$$

The problem of solving for the diffusion of trace elements in the basic model therefore reduces to the calculation of the inverse of the matrix $(\mathbf{I} - \mathbf{U})$ and the powers of the matrix $\mathbf{U}$. For example, since all entries in the column vector $[\mathbf{c}(a) - \mathbf{c}(\bar{a})]$ equal $-n/(N - 1)$ except the ath (which equals n), the decay of the trace after a single presentation of S_a is determined by the behavior of the diagonal entry u_{aa} of $\mathbf{U}$ as that matrix is raised to higher and higher powers. Indeed reasons will be adduced in the appendix for believing that the elements of the matrix $\mathbf{U}$ are always constrained in such a way that in general u_{aa}, and hence the trace of S_a, decreases with delay, d, like a sum of weighted exponential decay functions. Typically, of course, the exact values of the N^2 diffusion parameters, v_{ik}, are not initially known and, so, the powers of $\mathbf{U}$ can not be directly calculated. In order to fit the trace model to experimental data and interpret the parameters, therefore, it seems desirable to place some very stringent restrictions on the diffusion parameters. This is done in the next section, where two special cases of the basic model are treated in detail.

Applying the Model to Data

Special case 1. *Stimuli equally similar.* An instructive special case of the basic model is obtained by assuming that all of the stimuli in the domain under study are equally similar to each other (i.e., that $v_{ik} = v$ for $i \neq k$). In this case the elements of the matrix $\mathbf{U}$ are given by

$$(11) \qquad u_{ik} = \begin{cases} u(V + v) & \text{for } i = k \\ uv & \text{for } i \neq k, \end{cases}$$

where, for convenience here and in what follows, V is used to stand for $1 - Nv$. An induction on d then verifies that the elements of $\mathbf{U}^d$ are given by

$$(12) \qquad u_{ik}^{(d)} = \frac{1}{N} u^d (1 - V^d) + \begin{cases} (uV)^d & \text{for } i = k \\ 0 & \text{for } i \neq k. \end{cases}$$

From this, an expression can immediately be written for the elements of the matrix $(\mathbf{I} - \mathbf{U}^T)$. Finally, the elements of the matrix $(\mathbf{I} - \mathbf{U})^{-1}$ can be shown to be

$$(13) \qquad \frac{u(1 - V)}{N(1 - u)(1 - uV)} + \frac{1}{(1 - u)(1 - uV)} \begin{cases} (1 - u) & \text{for } i = k \\ 0 & \text{for } i \neq k, \end{cases}$$

since multiplication of this matrix by $(\mathbf{I} - \mathbf{U})$ then yields the identity matrix $\mathbf{I}$. These expressions, together with (6) and (9), can now be used to develop an explicit expression for any entry in the column vector $\mathbf{n}(T)$. In particular the number of trace elements associated with the representation for the dth from the last stimulus, namely S_a, is given by the ath entry in $\mathbf{n}(T)$. However, in order to exhibit the delay since the presentation of S_a explicitly, this entry will henceforth be denoted by $n_d(T)$. Equation (10) then yields for this entry

$$(14) \qquad n_d(T) = n \frac{1 - U^T}{N(1 - u)} - n \frac{1 - (uV)^T}{N(1 - uV)} + n(uV)^d.$$

According to the model, if there is no response bias, the probability of classification of a stimulus as "old" is proportional to the number of trace elements associated with its internal representation and, furthermore, becomes certainty if there is no delay. This implies that the unbiased probability of recognition of the dth from the last stimulus of the inspection sequence is given by

$$(15) \qquad P_d^*(T) = \frac{n_d(T)}{n_0(T)}.$$

Thus, regardless of the length of the inspection sequence (T),

$$(16) \qquad P_0^*(T) = 1,$$

as required.

By (1), then, the probability of the response "old" for any response bias B is given by

$$(17) \qquad\qquad P_d(T) = \left\{\frac{n_d(T)}{n_0(T)}\right\}^B.$$

If (14) is used to substitute for $n_d(T)$ and $n_0(T)$, the final result becomes, after simplification,

$$(18) \qquad P_d(T) = \left\{ 1 - \frac{1 - (uV)^d}{1 + \dfrac{1 - u^T}{N(1 - u)} - \dfrac{1 - (uV)^T}{N(1 - uV)}} \right\}^B.$$

This equation is an important consequence of the special case of the trace model in which all the stimuli are equally similar to each other (i.e., in which $v_{ik} = v$ for $i \neq k$). If T is held constant while d is varied, this equation describes a curve of forgetting. As can be seen, if the response bias is not too pronounced (i.e., if B is close to 1), the probability of correct recognition of a stimulus decreases like a simple exponential decay function of the delay since that stimulus was last presented. This special case of the trace model therefore leads to the same curve of forgetting as the earlier models proposed by London [13], von Foerster [21], Estes [5], Witte [24], and Bower [3]. In addition, however, this equation provides information about how the probability of falsely classifying a new stimulus as "old" depends upon T. For, if this probability of "false alarm" is denoted by $P_-(T)$, then

$$(19) \qquad\qquad P_-(T) = \lim_{d \to \infty} P_d(T).$$

(That is, since trace elements eventually disappear altogether, a stimulus that has not been presented for a sufficiently large number of trials is like a completely new stimulus.) Thus, as $d \to \infty$, (18) describes how the probability that a new stimulus will be classified as "old" increases with the total number of stimuli presented, T.

That the special case of the trace model just considered leads (like the earlier models) to a nearly exponential curve of forgetting can be regarded as a deficiency of that special case. For, as already noted, many empirically obtained curves seem to depart systematically from this predicted curve. Moreover, in some recent unpublished experiments, the departure could not be removed by varying the response bias, B. When such a discrepancy arises, it seems reasonable to consider the possibility that the deficiency stems primarily from the very restrictive assumption of the special case; namely, the assumption that all N stimuli are equally similar to each other. In fitting the trace model to data it is necessary, however, to keep the number of free parameters at a minimum. Under these conditions, perhaps the best strategy is to proceed to the next most complex case of the trace model.

Special case 2. *Two levels of similarity of the stimuli.* Specifically, for this second case of the trace model, the similarities between the stimuli will be assumed to take on either of two possible values. This can be accomplished by considering that the internal representations are of two kinds: isolated representations each of which corresponds to a relatively distinctive stimulus, and clustered representations each of which corresponds to a stimulus that is quite similar to several others. (In the domain of three-digit numbers, for example, the representation for 444 might be isolated while the representations for 335, 533, 355, and 553 might form a cluster.) Accordingly, there will be two rates of diffusion: a relatively low rate, v, between any two representations (isolated or clustered) so long as they are not members of the same cluster, and a higher rate, v_c , between any two representations that are members of the same cluster. If all clusters are considered to contain the same number of representations, N_c , then derivations like those used to obtain (18) yield

(20)

$$
P_d(T) = \left\{ 1 - \frac{1 - w(uV)^d - (1 - w)(uV_c)^d}{1 + \dfrac{1 - u^T}{N(1 - u)} - \dfrac{w[1 - (uV)^T]}{N(1 - uV)} - \dfrac{(1 - w)[1 - (uV_c)^T]}{N(1 - uV_c)}} \right\}^B
$$

where V_c stands for $V - N_c(v_c - v)$. An interpretation can also be given for the weight w: $(N - 1)w + 1$ is the number of isolated representations plus the number of clusters. (The number of isolated representations can, of course, be zero.)

If the response bias is not extreme in this second special case of the trace model, the curve of forgetting approximates a weighted sum of two exponential decay functions. Equation (20) is therefore consistent with the tentative conclusion (stated in the appendix) that, in general, the trace elements dissipate like a sum of exponentials. This result is encouraging because such a function always drops more rapidly at first and then more slowly than the best fitting simple exponential. As already remarked, this is just the way in which data often seem to depart from a simple exponential. One could proceed to derive equations like (18) and (20) for still more complex special cases in which three, four, etc., exponential functions appear. However, the utility of proceeding further in this direction is probably very small since the number of parameters to be estimated from the data would quickly become excessive. Hopefully, most empirical results could be adequately fitted without resorting to these still more complex cases. This does not mean that the stimuli are always grouped in clusters exactly as assumed in the derivation of (20). It merely means that, when the stimuli vary so greatly in similarity that the assumption of equal similarity fails, the assumption of clustering may nevertheless provide a satisfactory approximation.

Even if (20), say, should fit the data, the five parameters (u, V, V_e, w, and B) may seem too numerous. However, the one set of parameters should suffice to describe two curves: namely, $P_d(T)$ as a function of d, and $P_-(T)$ as a function of T. So, on the average, only $2\frac{1}{2}$ parameters are to be estimated from each empirical curve. Moreover, since the parameters have interpretations within the model, their existence should be independently demonstrable with different experimental operations (as, for example, in generalization experiments [18]). In this connection, u, V, V_e, and w are presumably functions primarily of the population of stimuli chosen. B, however, reflects the state of the subject and should be readily manipulable (e.g., by arranging different pay-off contingencies). Finally, when the model has been fitted to the data (regardless of the number of parameters used), it provides a rational basis for making the extrapolations required for the kind of informational estimate developed in the second part of this paper.

The Informational Analysis

An estimate will now be proposed for the total amount of information that is retained from the inspection sequence when that sequence is arbitrarily protracted, i.e., when $T \to \infty$. This estimate is to be calculated solely from the probability, $P_-(\infty)$, of false recognition together with the probabilities, $P_d(\infty)$, of correct recognition for all delays $0 \leq d < \infty$. Now these probabilities could in principle be estimated empirically by administering sufficiently long inspection sequences. In practice, however, the length of the inspection sequence required would generally be prohibitive. Accordingly, some basis is needed for extrapolating from the directly estimated probabilities $P_-(T)$ and $P_d(T)$ with $0 \leq d < T$, for some practicable value of T, to the larger set of asymptotic probabilities $P_-(\infty)$ and $P_d(\infty)$ with $0 \leq d < \infty$. The trace model developed in the first part of this paper will be used as the basis for this extrapolation. The informational analysis to be developed here does not, however, rest in any essential way upon the specific assumptions of the trace model. Thus, even if the trace model should subsequently be discarded in favor of some improved model for the memory process, the informational analysis would presumably still be applicable.

Suppose the model has been fitted to actual data, and an equation has been found relating $P_d(\infty)$ to d. Owing to the asymptotic behavior of $P_d(\infty)$, the subject's span of retention is roughly limited by a number M such that

$$(21) \qquad P_d(T) \cong \begin{cases} P_-(T) & \text{for} \quad d \geq M \\ P_d(\infty) & \text{for} \quad T > M. \end{cases}$$

Thus, if $T \geq M$, the subject is effective in a steady state and his behavior is essentially characterized by the $M + 1$ probabilities

$$(22) \qquad P_\infty(\infty) \quad \text{and} \quad P_d(\infty) \quad \text{with} \quad 0 \leq d < M.$$

Consider, then, an observer who knows two things: (a) the set of $M + 1$ probabilities (22) that characterize a given subject; and (b) that in a given experiment this subject attained steady state, i.e., that $T \geq M$. Let $H(R)$ denote the uncertainty of this observer as to which of the N stimuli would be classified by the subject as "old" under the condition that this observer has no knowledge about which inspection sequence was in fact presented. Similarly, let $H_s(R)$ denote the uncertainty of this observer as to which of the N stimuli would be classified as "old" under the condition that this observer knows exactly which inspection sequence was presented. Now, the total amount of information about the inspection sequence that is retained by the subject cannot be less than the reduction in the observer's uncertainty (about which stimuli would be classified as "old") that would result solely from a knowledge of what inspection sequence was in fact presented; that is, it cannot be less than

$$(23) \qquad\qquad H(R) - H_s(R)$$

([16], p. 38). Thus the desired informational estimate is obtained if expressions can be developed for $H(R)$ and $H_s(R)$ in terms of the probabilities (22).

Before proceeding to this development, however, certain preliminary points should be clarified. First, the information to be estimated here is the total amount retained by the subject immediately following the removal of the last stimulus in the inspection sequence; this amount may considerably exceed the amount that the subject could actually transmit to an observer by responding to any given recognition array. For, as already indicated, the memory traces would largely dissipate before the subject could respond to each stimulus in an array of size N. Still, it is legitimate to estimate the total amount retained by testing for recognition of just a small sample of the N stimuli—say those corresponding to certain representative delays, d. This is analogous to estimating how much a student has retained from a course by testing him on a small sample of items from the course. A similar sampling method has also been recently employed by Sperling to obtain informational estimates for short-term memory [19]. The second point is that the estimate obtained from (23) is strictly only a lower bound. This is because (contrary to the model) it is possible that the subject retains some information about the order of presentation of stimuli that is not represented in his responses of "old" or "new," and because (if the domain of stimuli is not properly chosen) the size of the effective set of stimuli from the standpoint of the subject may somewhat exceed the size N of the experimenter's population of stimuli.

With these provisos in mind, expressions will now be developed for the terms in (23). Since the probabilities $P_d(\infty)$ are for independent events, the second term is given by

$$(24) \quad H_S(R) = - \sum_{d=0}^{M-1} \{P_d \log_2 P_d + (1 - P_d) \log_2 (1 - P_d)\}$$

$$- (N - M)\{P_\infty \log_2 P_\infty + (1 - P_\infty) \log_2 (1 - P_\infty)\},$$

where, since steady state is understood, P_d is written for $P_d(\infty)$. Furthermore, if $Q(\nu)$ denotes the probability that the subject would classify exactly ν out of the N stimuli as "old," then the a priori probability that any particular subset of stimuli would be so classified is $Q(\nu)/\binom{N}{\nu}$ and, hence,

$$(25) \quad H(R) = - \sum_{\nu=0}^{N} Q(\nu) \log_2 \frac{Q(\nu)}{\binom{N}{\nu}}.$$

Expressions (23) through (25) yield the results that seem intuitively correct for certain simple cases. For example, if the curve of forgetting were a step function such that

$$P_d(\infty) = \begin{cases} 1 & \text{for} \quad d < M \\ 0 & \text{for} \quad d \geq M, \end{cases}$$

then

$$Q(\nu) = \begin{cases} 1 & \text{for} \quad \nu = M \\ 0 & \text{otherwise,} \end{cases}$$

and

$$H(R) - H_S(R) = \log_2 \binom{N}{M}.$$

That is, the subject would then be carrying at least the $\log_2 \binom{N}{M}$ bits that specify which subset of M out of the N stimuli constituted the last M stimuli of the inspection sequence.

For the kinds of curves of forgetting prescribed by the model or obtained empirically, however, the number of steps in the direct calculation of the probabilities $Q(\nu)$ becomes prohibitive. Fortunately, a normal (Gaussian) approximation to the distribution of the $Q(\nu)$ serves quite well. In order to show this, $Q_d(\nu)$ will be used to denote the probability that exactly ν out of the d last stimuli of the inspection sequence would be classified as "old." Thus

$$Q(\nu) = \lim_{d \to \infty} Q_d(\nu).$$

The probabilities $Q_d(\nu)$ were actually computed for $1 \leq d \leq 20$ from the

probabilities $P_d(\infty)$ obtained by fitting the model to data from a recognition experiment using three-digit numbers as stimuli. For each d-value the calculated distribution of the $Q_d(\nu)$ was compared with the normal distribution with mean

$$(26) \qquad \mu_d = \sum_{i=0}^{d-1} P_i(\infty)$$

and variance

$$(27) \qquad \sigma_d^2 = \sum_{i=0}^{d-1} P_i(\infty)[1 - P_i(\infty)].$$

The fit of the normal curve systematically improved as d increased. As seen in Fig. 2, the degree of approximation for $d = 20$ was quite close. This approximation will probably be satisfactory as long as the curve of forgetting does not drop too precipitously (cf., [7], pp. 9 and 125).
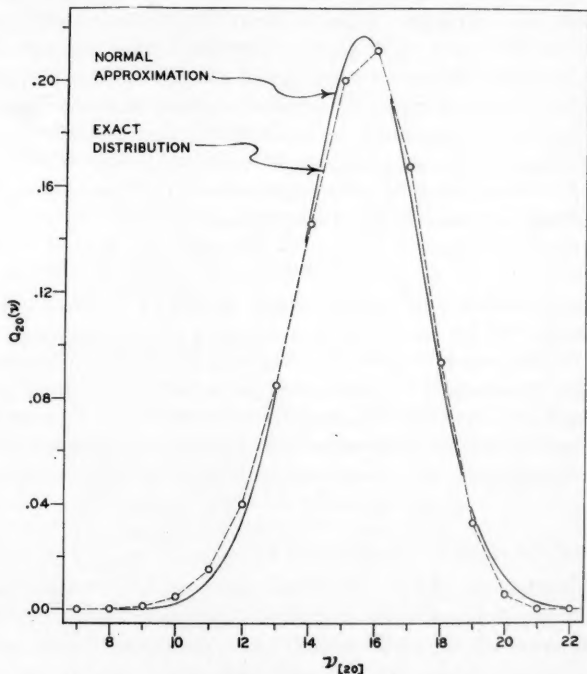

FIGURE 2

The Exact Distribution and the Normal Approximation of $Q_{20}(\nu)$, the Probability of Classifying as "Old" ν of the Final 20 Stimuli

If the normal curve is taken to approximate the distribution of the probabilities $Q(\nu)$, then, Shannon's formula for the uncertainty associated with a normal distribution ([16], p. 56) can be used to rewrite (25) as follows:

$$(28) \quad H(R) \cong - \sum_{\nu=0}^{N} Q(\nu) \log_2 Q(\nu) + \sum_{\nu=0}^{N} Q(\nu) \log_2 \binom{N}{\nu}$$

$$= \log_2 \sqrt{2\pi e}\sigma_N + \log_2 N!$$

$$- \frac{1}{\sqrt{2\pi}\sigma_N} \sum_{\nu=0}^{N} \exp\left[-\left(\frac{\nu - \mu_N}{\sqrt{2}\sigma_N}\right)^2 \right] \cdot [\log_2 (\nu!) + \log_2 [(N - \nu)!]].$$

Thus $H(R)$ and, hence, $H(R) - H_s(R)$ can be estimated from the empirically determined probabilities of recognition. For example, when three-digit numbers were used as stimuli, (28) led to an estimate of 32 bits for the total amount of information carried along by a subject in steady state.

Actually, according to the model, the probabilities upon which the informational calculations are based may be biased. Furthermore, the estimated amount of information depends upon the bias, as is clear from the fact that it vanishes as $B \to 0$. Thus, to the extent that the trace model is accepted, it furnishes a way of maximizing the estimated lower bound on the retained information through variation of the parameter B. However, estimates obtained by varying B away from its experimentally determined value use the model as more than simply a rational basis for extrapolating curves. Such estimates rest more heavily on the specific assumptions of the model and, hence, are necessarily much more tentative.

Appendix

Two kinds of elaborations that might eventually be incorporated into the trace model will be indicated. One concerns a specification of the kind of constraints that must in general be imposed on the diffusion parameters, v_{ik}, and the consequences of these constraints for the general form of the curve of forgetting. The other concerns the possibility of replacing the somewhat arbitrary power transformation assumed for the response bias by a set of more compelling assumptions about the underlying micromechanical process.

The General Form of the Curve of Forgetting

According to (10), the way in which the trace elements dissipate after the presentation of a particular stimulus S_a is determined by the behavior of the diagonal entry u_{aa}^{d} of $\mathbf{U}^d = (u\mathbf{V})^d$ with increasing d. But, unless the v_{ik} are constrained in some way, this entry can exhibit periodicities of a kind that have not been found in actual curves of forgetting. Presumably, then, the v_{ik} must be subject to some general constraints. In the earlier application of the trace model it was proposed that the rate of diffusion between

two representations was a decreasing function of the distance between these representations in their "psychological space" [18]. The geometry of this space, according to this view, is what constrains the v_{ik}. At the very least, for example, the v_{ik} must be consonant with the metric axioms for the distances between the representations [17], namely,

$$D_{ii} = 0, \qquad D_{ik} = D_{ki}, \qquad D_{ik} \leq D_{ii} + D_{ik}.$$

With a proper choice of the monotonic function relating diffusion to distance, Assumption I of the trace process is already consonant with the first two axioms for $v_{ii} \geq v_{ik} = v_{ki}$. These conditions, together with the triangle inequality (the third metric axiom), imply that $\mathbf{U}$ is real, symmetric, and irreducible. The consequences of these constraints is that there exists a real orthogonal matrix, $\mathbf{G}$, such that

$$\mathbf{U}^d = \mathbf{G}\mathbf{\Lambda}^d\mathbf{G}',$$

where $\mathbf{\Lambda}$ is the diagonal matrix containing the characteristic roots λ_i of $\mathbf{U}$, and where these roots are real with max $|\lambda_i| = u < 1$ ([6], vol. 1, p. 308, and vol. 2, pp. 53 and 63). From this it follows that the subsequence of entries $u_{aa}^{(d)}$ with even superscripts, d, decreases like a sum of exponential decay functions and, hence, is completely monotonic ([22], p. 108). The entire sequence may still exhibit fluctuations of period 2. Nevertheless, for spaces with certain plausible metrics (e.g., Euclidean) and for certain reasonable functions relating the v_{ik} to the D_{ik}, $\mathbf{U}$ can be shown to be positive semidefinite. Since in this case none of the λ_i can be negative, the entire sequence is completely monotonic and, in fact, decreases like a sum of exponentials. This, then, is the basis for the conjecture that curves of forgetting can always be fitted by a sum of exponential decay functions.

Possible Reformulations of the Response Bias

One way in which the power transformation (1) in Assumption V might be replaced by a more rational mechanism for the response bias is suggested by Egan's recent application of the theory of signal detectability to recognition memory [4]. The following modifications of the trace model might suffice. (*a*) The number n of trace elements introduced on each trial (instead of being large, as assumed above) would have to be small enough that statistical fluctuations in the number of elements transferring from one representation to another become substantial. The number of elements associated with a particular representation on a given trial would thus have not a fixed value but, rather, any of the several values specified by some distribution. (*b*) The subject would then be conceived as establishing a given bias by selecting a criterion cut with respect to the number of trace elements that must be associated with an internal representation before he will classify the correspondng stimulus as "old." A similar notion has been suggested by

J. R. Pierce (personal communication). He pointed out that, if the trace elements are regarded as moving at random in a space, the probability of finding at least one element in a given region (e.g., in the region S_a^* corresponding to the stimulus S_a) is of the form

$$P_a = 1 - e^{-n_a},$$

where n_a is the average density of trace elements in the vicinity of the given region. This function is convex upward and, indeed, was not very different from the power transformation of Fig. 1 for the positive bias found in the experiment with three-digit numbers. With this formulation, the amount of bias could be manipulated either by varying the area included in the region for each stimulus or, again, by varying the number of elements that must fall in that region.

REFERENCES

[1] Bachem, A. Time factors in relative and absolute pitch determination. *J. acoust. Soc. Amer.*, 1954, **26**, 751–753.

[2] Baldwin, J. M. and Shaw, W. J. Memory for square-size. *Psychol. Rev.*, 1895, **2**, 236–239.

[3] Bower, G. H. A theory of serial discrimination learning. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford, Calif.: Stanford Univ. Press, 1959. Pp. 76–93.

[4] Egan, J. P. Recognition memory and the operating characteristic. Tech. Note AFCRC-TN-58-51, AD-152650, Indiana Univ.: Hearing and Communication Lab., 1958.

[5] Estes, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.

[6] Gantmacher, F. R. *The theory of matrices*. New York: Chelsea, 1959. 2 vols.

[7] Gnedenko, B. V. and Kolmogorov, A. N. *Limit distributions for sums of independent random variables*. Cambridge, Mass.: Addison-Wesley, 1954.

[8] Goldberg, S. *Introduction to difference equations*. New York: Wiley, 1958.

[9] Hanawalt, N. G. Memory trace for figures in recall and recognition. *Arch. Psychol. N. Y.*, 1937, No. 216.

[10] Harris, J. D. The decline of pitch discrimination with time. *J. exp. Psychol.*, 1952, **43**, 96–99.

[11] Jenkins, J. G. and Dallenbach, K. M. Oblivescence during sleep and waking. *Amer. J. Psychol.*, 1924, **35**, 605–612.

[12] Leyzorek, M. Two-point discrimination in visual space as a function of the temporal interval between the stimuli. *J. exp. Psychol.*, 1951, **41**, 364–375.

[13] London, I. D. An ideal equation for a class of forgetting curves. *Psychol. Rev.*, 1950, **57**, 295–302.

[14] Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.

[15] Luh, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, No. 142.

[16] Shannon, C. E. and Weaver, W. *The mathematical theory of communication*. Urbana: Univ. Illinois Press, 1949.

[17] Shepard, R. N. Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, **22**, 325–345.

[18] Shepard, R. N. Stimulus and response generalization: deduction of the generalization gradient from a trace model. *Psychol. Rev.*, 1958, **65**, 242–256.

[19] Sperling, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**, No. 11 (whole No. 498).

[20] Strong, E. K. The effect of time-interval upon recognition memory. *Psychol. Rev.*, 1913, **20**, 339–372.

[21] von Foerster, H. Quantum mechanical theory of memory. In H. von Foerster (Ed.), *Cybernetics, transactions of the sixth conference.* New York: Josiah Macy, Jr. Foundation, 1950.

[22] Widder, D. V. *The Laplace transform.* Princeton: Princeton Univ. Press, 1941.

[23] Williams, O. A study of the phenomenon of reminiscence. *J. exp. Psychol.*, 1926, **9**, 368–387.

[24] Witte, R. S. A stimulus-trace hypothesis for statistical learning theory. *J. exp. Psychol.*, 1959, **57**, 273–283.
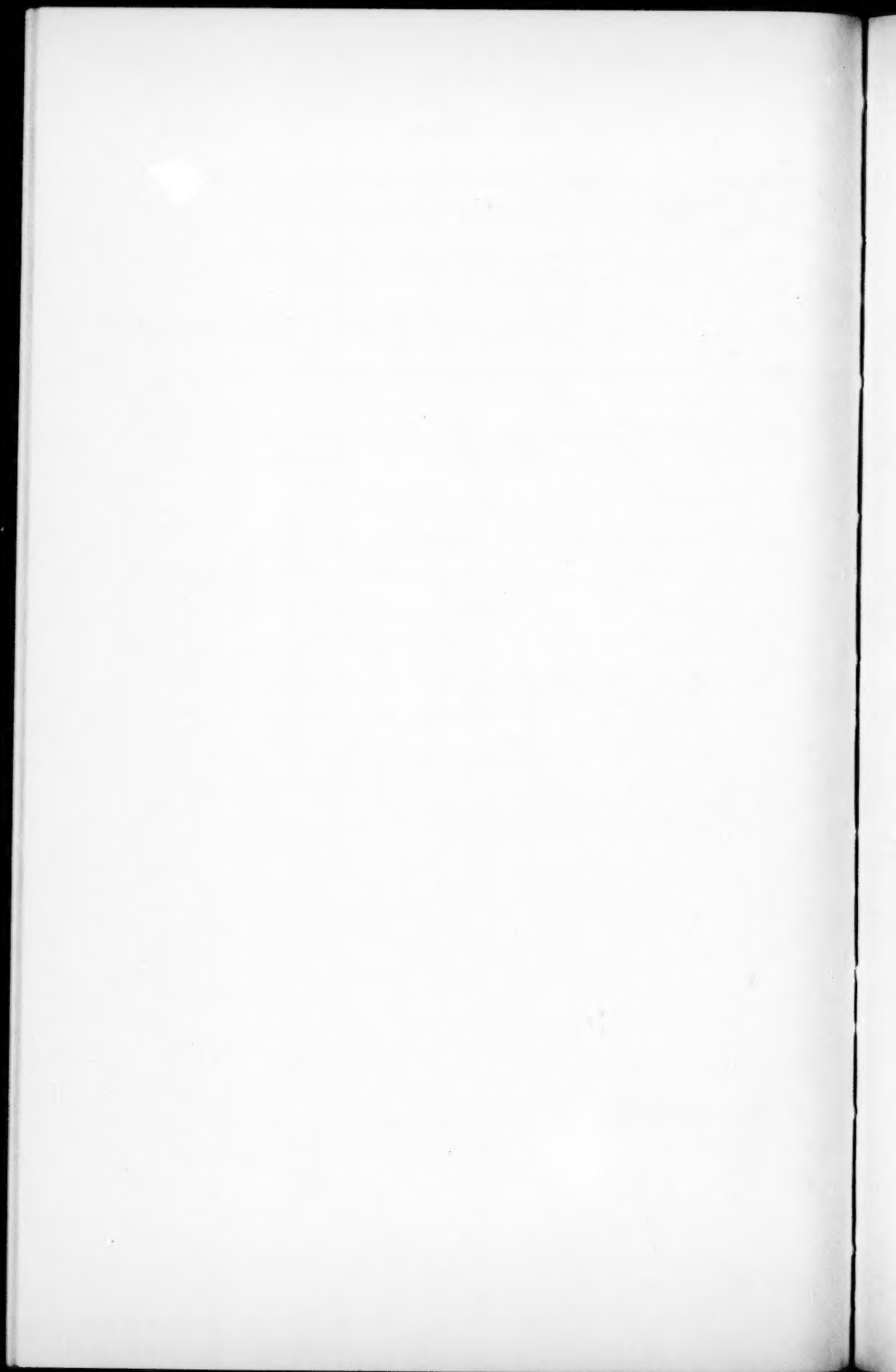
ANALYSIS OF UNREPLICATED THREE-WAY CLASSIFICATIONS, WITH APPLICATIONS TO RATER BIAS AND TRAIT INDEPENDENCE*

JULIAN C. STANLEY

UNIVERSITY OF WISCONSIN

The seven analysis-of-variance mean squares for an unreplicated three-way classification may be written as linear combinations of a mean variance and three mean covariances. Formulas are presented for computing the mean variances and mean covariances from linear combinations of mean squares. The relevance of these formulas for assessing rater biases and trait independence is discussed, a numerical example is provided, and proposed extensions are briefly noted.

When repeated measurements of individuals are made over all levels of two experimental variables, three sources of covariance become possible. Consider the familiar situation where each individual is rated once by each rater on each trait, there being at least two individuals, two raters, and two traits. Covariation can occur within each rater across traits, within each trait across raters, and across both raters and traits.

These three sources of covariation are orthogonal. Empirically, it has been found that covariation within raters across traits, inflated by relative halo effect, usually exceeds covariation within traits across raters, the magnitude of which reflects independence of the traits. Covariation across both raters and traits constitutes a baseline against which the other two sources may be judged. It tends to be less than either of them.

In 1954, Guilford ([8], p. 281) showed that the various rater biases can be thought of appropriately in terms of analysis-of-variance mean squares involving raters: the mean squares for (i) raters, (ii) the interaction of raters with ratees, (iii) the interaction of raters with traits, and (iv) the second-order interaction of raters with both ratees and traits. Thus, there are four possible sources of rater bias, three of which (the main effect of raters and the two first-order interactions) may be evaluated in a given study and compensated for statistically, as will be shown in this article.

The ratee-rater-trait matrix is used above merely as an introductory illustration. Also, ratees define rows only for convenience of initial exposi-

tion. Raters or traits might just as well define rows. In each instance, three sources of covariation can be isolated. Over the three orderings, each first-order interaction will be defined twice and the second-order interaction three times. In a generalized, complete three-classification matrix each interaction mean square may be shown to be a linear function of a mean within-column variance and three mean covariances among "levels" of the two factors defining columns.

Method of Analysis

Consider any matrix of real numbers X_{irt} , where $i = 1, 2, \cdots , I$; $r = 1, 2, \cdots , R$; and $t = 1, 2, \cdots , T$. Partition the total sum of squared deviations around the mean of these $I \times R \times T$ numbers into the usual seven sums of squares: three for main effects, three for first-order (two-factor) interactions, and one for second-order (three-factor) interaction.

The four mean squares (i.e., sums of squares divided by the appropriate number of degrees of freedom) involving i may be written

$$(1) \quad \mathrm{MS}_i \quad = A + (R - 1)B + (T - 1)C + (R - 1)(T - 1)D,$$

$$(2) \quad \mathrm{MS}_{(i \times r)} = A - \qquad B + (T - 1)C - \qquad (T - 1)D,$$

$$(3) \quad \mathrm{MS}_{(i \times t)} = A + (R - 1)B - \qquad C - (R - 1) \qquad D,$$

$$(4) \quad \mathrm{MS}_{(i \times r \times t)} = A - \qquad B - \qquad C + \qquad D,$$

where

$$A = \overline{s_{rt}^2} , B = \overline{\mathrm{cov}\,(X_{rt} , X_{r't})}, \; C = \overline{\mathrm{cov}\,(X_{rt} , X_{rt'})}, \; D = \overline{\mathrm{cov}\,(X_{rt} , X_{r't'})}$$

with $r \neq r'$ and $t \neq t'$. Bars denote means. (For an outline of the way in which the formulas were obtained, see the Appendix at the end of this paper.)

If, for convenience, the i factor is considered to define rows of the matrix and the other two factors columns, A is the mean of the RT within-column variances of the form

$$s_{rt}^2 = \sum_{i=1}^{I} (X_{irt} - \bar{X}_{.rt})^2 / (I - 1).$$

B is the mean of the $T[R(R - 1)]$ covariances across the R raters within the T traits. C is the mean of the $R[T(T - 1)]$ covariances across t's within r's. D is the mean of the remaining $RT(RT - 1) - RT(R - 1) - RT(T - 1) = RT(R - 1)(T - 1)$ covariances, those across both r's and t's.

Formulas (1)–(4), independent linear equations in four unknowns, can be solved for the mean variance and the three mean covariances to secure the following formulas, where $MS_i = w$, $\mathrm{MS}_{(i \times r)} = x$, $\mathrm{MS}_{(i \times t)} = y$, and $\mathrm{MS}_{(i \times r \times t)} = z$.

$$(5) \quad A = [w + (R - 1)x + (T - 1)y + (R - 1)(T - 1)z]/RT,$$

(6) $B = [w - \qquad x + (T - 1)y - \qquad (T - 1)z]/RT,$

(7) $C = [w + (R - 1)x - \qquad y - (R - 1) \qquad z]/RT,$

(8) $D = [w - \qquad x - \qquad y + \qquad z]/RT.$

By treating factor r as defining rows and factors i and t as defining columns, one obtains expressions analogous to those of (1)–(4):

(9) $\mathrm{MS}_r \qquad = E + (I - 1)F + (T - 1)G + (I - 1)(T - 1)H,$

(10) $\mathrm{MS}_{(r \times i)} = E - \qquad F + (T - 1)G - \qquad (T - 1)H,$

(11) $\mathrm{MS}_{(r \times t)} = E + (I - 1)F - \qquad G - (I - 1) \qquad H,$

(12) $\mathrm{MS}_{(r \times i \times t)} = E - \qquad F - \qquad G + \qquad H,$

where

$$E = \overline{s_{it}^2}, \qquad\qquad F = \overline{\mathrm{cov}\,(X_{it}\,,\,X_{i't})},$$
$$G = \overline{\mathrm{cov}\,(X_{it}\,,\,X_{it'})}, \qquad H = \overline{\mathrm{cov}\,(X_{it}\,,\,X_{i't'})},$$

with $i \neq i'$ and $t \neq t'$.

Solving (9)–(12), one obtains the following formulas, where

$$\mathrm{MS}_r = u, \qquad \mathrm{MS}_{(r \times i)} = \mathrm{MS}_{(i \times r)} = x, \qquad \mathrm{MS}_{(r \times t)} = v,$$
$$\mathrm{MS}_{(r \times i \times t)} = \mathrm{MS}_{(i \times r \times t)} = z.$$

(13) $E = [u + (I - 1)x + (T - 1)v + (R - 1)(T - 1)z]/IT,$

(14) $F = [(T - 1)(v - z) + u - x]/IT,$

(15) $G = [(I - 1)(x - z) + u - v]/IT,$

(16) $H = [(u - x - v + z)]/IT.$

Finally, treating t as defining rows and i and r as defining columns,

(17) $\mathrm{MS}_t \qquad = J + (I - 1)K + (R - 1)L + (I - 1)(R - 1)M,$

(18) $\mathrm{MS}_{(t \times i)} = J - \qquad K + (R - 1)L - \qquad (R - 1)M,$

(19) $\mathrm{MS}_{(t \times r)} = J + (I - 1)K - \qquad L - (I - 1) \qquad M,$

(20) $\mathrm{MS}_{(t \times i \times r)} = J - \qquad K - \qquad L + \qquad M,$

where

$$J = \overline{s_{ir}^2}, \qquad\qquad K = \overline{\mathrm{cov}\,(X_{ir}\,,\,X_{i'r})},$$
$$L = \overline{\mathrm{cov}\,(X_{ir}\,,\,X_{ir'})}, \qquad M = \overline{\mathrm{cov}\,(X_{ir}\,,\,X_{i'r'})},$$

with $i \neq i'$ and $r \neq r'$. Solving (17)–(20) one obtains the following formulas,

where

$$\mathrm{MS}_t = q, \quad \mathrm{MS}_{(t\times i)} = y, \quad \mathrm{MS}_{(t\times r)} = v, \text{ and } \mathrm{MS}_{(t\times i\times r)} = z.$$

(21) $J = [q + (I - 1)y + (R - 1)v + (I - 1)(R - 1)z]/IR,$

(22) $K = [(R - 1)(v - z) + q - y]/IR,$

(23) $L = [(I - 1)(y - z) + q - v]/IR,$

(24) $M = [(q - y - v + z)]/IR.$

Note that each of the two-factor interactions is defined twice while the three-factor interaction is defined thrice. For example, by (2) and (10),

$$\mathrm{MS}_{(i\times r)} = A - B + (T - 1)C - (T - 1)D$$
$$= E - F + (T - 1)G - (T - 1)H.$$

By (4) and (12),

$$\mathrm{MS}_{(i\times r\times t)} = A - B - C + D = E - F - G + H.$$

Ignoring the mean squares themselves and subtracting the expressions for $\mathrm{MS}_{(i\times r)}$ from corresponding expressions for $\mathrm{MS}_{(i\times r\times t)}$,

$$C - D = G - H.$$

In other words,

$$\overline{\mathrm{cov}\,(X_{rt}\,,\,X_{rt'})} - \overline{\mathrm{cov}\,(X_{rt}\,,\,X_{r't'})} = \overline{\mathrm{cov}\,(X_{it}\,,\,X_{it'})} - \overline{\mathrm{cov}\,(X_{it}\,,\,X_{i't'})}.$$

Similarly,

$$B - D = L - M, \text{ and } F - H = K - M.$$

Note in particular the following relationships:

(25) $B - D = L - M = \dfrac{1}{R}\,(y - z) = \dfrac{1}{R}\,[\mathrm{MS}_{(i\times t)} - \mathrm{MS}_{(i\times r\times t)}],$

(26) $C - D = G - H = \dfrac{1}{T}\,(x - z) = \dfrac{1}{T}\,[\mathrm{MS}_{(i\times r)} - \mathrm{MS}_{(i\times r\times t)}],$

and

(27) $F - H = K - M = \dfrac{1}{I}\,(v - z) = \dfrac{1}{I}\,[\mathrm{MS}_{(r\times t)} - \mathrm{MS}_{(i\times r\times t)}].$

Multirater-Multitrait Matrices

The above formulas (no significance tests implied) pertain to any complete matrix of real numbers, however gathered and regardless of what i, r, and t happen to represent. An especially important application occurs when

i designates ratees, r designates raters, and t designates traits. From the work of Guilford [8], Willingham and Jones [19], and others, the three mean squares involving ratees (r)—MS_r , $MS_{(i \times r)}$, and $MS_{(r \times t)}$—may reflect, respectively, differences among some raters in general level of rating, bias of some raters toward certain individuals, and bias of some raters toward certain traits. Finally, $MS_{(i \times t)}$ reflects differential meaning of the various traits, as Willingham and Jones [19] have shown. "Valid variance" in ratee-rater-trait studies usually consists chiefly of the variance components σ_i^2 and $\sigma_{(i \times t)}^2$; sometimes differences in trait means may be of interest, too.

From the definitions of B, C, and D, one sees that B is the mean of the covariance within traits ($t = t$) across raters ($r \neq r'$). For the tth trait there are $R(R - 1)$ covariances among the R raters, half of these being duplicates of the other half because cov $(r, r') =$ cov (r', r). C is the mean of the covariances within raters ($r = r$) across traits ($t \neq t'$). For the rth rater there are $T(T - 1)$ covariances among the T traits, half of them duplicates of the other half. D, on the other hand, is the mean of the covariances across both raters ($r \neq r'$) and traits ($t \neq t'$). Its magnitude reflects neither interaction of ratees with traits, as does B, nor bias of raters toward ratees, as does C. D constitutes the only internal base for evaluating the magnitudes of B and C. Typically, $B \geq D$ and $C \geq D$, though of course D could exceed B or C. In order to maximize differential meaning of the traits used, B should be as large as possible relative to D, and to minimize the bias of some raters toward certain ratees, $C \leq D$.

Chi ([3], p. 237) sensed part of this latter relationship when he wrote ". . . the correlation between two traits, according to the ratings by one rater, tends to be higher than it should be. On the other hand, since two raters are not likely to take the same attitude or to be under the same prejudice toward an individual rated, the correlation between two traits, according to the ratings by two different raters, would be relatively free from the halo effect. Hence the difference between the former and the latter correlation coefficients may be regarded as the halo effect contained in the ratings by one rater." He performed a factor analysis of such differences and found a general factor of halo, independent of the general factor of the ratings themselves, that accounted for about half as much of the total variance (17 vs. 32 percent).

In a given study, one may find any degree of relative halo effect and any degree of trait independence, for $MS_{(i \times r)}$ is independent of $MS_{(i \times t)}$. These constitute two *separate* criteria for the adequacy of ratings, as Campbell and Fiske [2] and Humphreys [12] point out with respect to multitrait-multimethod matrices. If one reads *rater* in the present paper for *method* in theirs, he has at his command some of the objective summary statistical procedures for which Campbell and Fiske asked.

Formula (25) shows that the difference between B and D (or L and M)

is a simple function of $\mathrm{MS}_{(i \times t)}$ and $\mathrm{MS}_{(i \times r \times t)}$. B reflects covariation common to all rt, $r't'$ pairings, plus covariation among the rt, $r't$ pairings. $B - D$ is estimated by the final part of (25). For example, (3) may be written as

$$(28) \qquad \mathrm{MS}_{(i \times t)} = (A - B - C + D) + R(B - D)$$

$$= \mathrm{MS}_{(i \times r \times t)} + R(B - D).$$

The "Pigeonhole" Model

How is one to get tests of significance for $B - D$, $C - D$, and $F - H$? From (28), the ratio $\mathrm{MS}_{(i \times t)} / \mathrm{MS}_{(i \times r \times t)}$ resembles an F ratio, with $(B - D)$ being the effect tested. Is this ratio in fact distributed as F under the null hypothesis? Less stringently, is the right-hand member of (25) an unbiased estimate of the variance component (using that expression in a broad sense) $\sigma^2_{(i \times t)}$? The answer to this latter question would seem to depend upon which analysis-of-variance model yields appropriate expected mean squares for the particular study conducted.

Consider the relatively unrestrictive general linear model set up by Cornfield and Tukey [4] for their "pigeonhole" model (which may also be generalized to an urn-sampling model). By extension of their model for two crossed factors ([4], p. 920), for the rating x_{irts} received by the ith ratee from the rth rater on the tth trait the sth time he is rated by that rater on that trait

$$x_{irts} = \theta + \alpha_i + \gamma_r + \delta_t + \epsilon_{ir} + \eta_{it} + \kappa_{rt} + \lambda_{irt} + \omega_{irts}.$$

Theta represents the general contribution, estimated by $\bar{X}_{\dots}$ (for the pigeonhole model, $\sigma^2_\theta = 0$). The next seven Greek letters denote the three main contributions and four interactions that are possible. Assumptions are as listed in ([4], pp. 920–921).

Expected mean squares for the finite case of the above linear model are given in ([4], p. 929). (For rather similar $E[\mathrm{MS}]$'s, see [14].) Under what conditions do formulas (25)–(27) estimate variance components without bias?

The right-hand member of (25) estimates the variance component $\sigma^2_{(i \times t)}$ if the raters used in the study were drawn randomly from a large population of raters. The right-hand member of (26) estimates $\sigma^2_{(i \times r)}$ if the traits used in the study were drawn randomly from a large population of traits. The right-hand member of (27) estimates $\sigma^2_{(r \times t)}$ if the ratees used in the study were drawn randomly from a large population of ratees. Otherwise, the respective variance components will tend to be underestimated by formulas (25)–(27), unless $\sigma^2_{(i \times r \times t)} = 0$, as will the analogous F-ratios computed to test significance.

Usually, investigators capture "grab groups" of ratees and raters, who

then constitute the entire population "sampled." Such groups may be composed of volunteers or entire "handcuffed volunteer" classes, but rarely are individuals (ratees or raters) sampled randomly from any defined population. In view of the three conclusions reached above concerning variance components and tests of significance, this appears disturbing. Nearly always we want to generalize beyond the particular ratees and raters used in the study to other ratees and raters "like them." In repeating the study, we would probably use new ratees and raters, but the same traits (though in a given study we might have each ratee rated more than once by each rater on each trait, as allowed for in the above model).

Can we merely *consider* the ratees used in the study as a random sample from a large hypothetical population of ratees "like themselves," and consider the raters similarly? If so, we would have a mixed model (ratees and raters random, traits fixed) for which (25) and (27) would yield unbiased estimates of the variance components $\sigma^2_{(i \times t)}$ and $\sigma^2_{(r \times t)}$.

Cornfield and Tukey ([4], pp. 913–914) tend to encourage this "bootstrap randomization," while Wilk and Kempthorne ([16], pp. 1162–1163; [18], pp. 953–954) discourage it. The latter writers remark: "There are some circumstances under which it may be useful to consider the levels of a random factor actually used as though they were the levels of a fixed factor (with a corresponding redefinition of main effects and interactions), but there appears to be no objective basis for the converse case" ([16], p. 1163).

The matter seems by no means settled yet. By adopting the Cornfield-Tukey point of view we are of course "better off" with the unreplicated ratees-raters-traits study than we would be under the greater restrictions of the Wilk-Kempthorne approach. Replication seems desirable in most instances, however, both within ratee-rater-trait "cells" and across studies with other "grab groups" of ratees and raters. It may be best to assume a fixed-effects model and use $\mathrm{MS}_{(i \times r \times s \times t)}$ for testing all effects and interactions in a given study.

Replication within a given study has the added advantage of revealing further biases of raters: $i \times r \times t, i \times r \times s, r \times s \times t,$ and $r \times s$. These can be compensated for statistically in a manner analogous to that of (30), which appears later in this article.

Trait Independence and Rater Bias

B/A should be a close estimate of $\bar{r}_{rt,r't}$, the mean correlation among raters within traits. D/A should be a close estimate of $\bar{r}_{rt,r't'}$, the mean correlation across both raters and traits.

If B significantly exceeds D, then it may be worthwhile to weight the trait scores differentially for predictive purposes. If it does not, then the standard score of the ith individual differs only randomly from trait to trait, and differential weighting is futile. (Here, for the fixed-effects model, we

assume again that $MS_{(i \times r \times t)}$ has as its expected value σ_E^2 , pure error-of-measurement variation [4].)

When statistical significance occurs for $MS_{(i \times t)}$, one may want to find a linear combination of trait factor scores that maximizes the ratio $MS_i / MS_{(i \times r)}$, thereby making differences among the means of individuals as large as possible relative to rater bias toward individuals. This is one way to correct for what Guilford ([8], p. 284) calls relative halo effects. Abelson [1] shows how to employ linear discriminant analysis to maximize variance ratios of this sort. Bias of raters toward ratees is usually so strong that in large studies the $i \times r$ interaction probably shows up as significant, even when $MS_{(i \times r \times t)}$ is used as the error term. Independence of traits and biases of raters toward traits seem less potent.

The better controlled the investigation, the closer D/A will probably approach zero—that is, the poorer the correlation across both raters and traits. (There is, of course, the problem of generally prejudicing information, affecting several raters across traits within ratees.) Careful randomization of the order of presentation of the $I \times T$ ratee-trait combinations, independently for each rater, when experimentally feasible might reduce the extent of interactive rater biases and perhaps increase the independence of traits. (Johnson and Vidulich [13] tried two orders, all traits for one individual vs. all individuals for one trait, but apparently did not randomize anything.)

Consideration of the various possibilities for randomizing the order of ratees, raters, and/or traits used, and of their influences upon expected mean squares, is beyond the scope of this paper; suffice it to say that the analysis of variance mentioned above presupposes *complete* randomization of the order of the $I \times R \times T$ combinations. Kempthorne and collaborators, having contributed greatly to analysis of completely and restrictively randomized designs [16, 17, 18], are now devising analyses (structures) for situations where randomization within the experiment itself can vary from little or none to much or complete, as in the ratee-rater-trait type of investigation. Generally, expected mean squares are considered by them to depend upon what randomization actually takes place within the study (this in addition to the sampling of levels of the factors themselves).

Probably we are well advised to design fuller studies, in which each rater rates each ratee at least twice on each trait. Then there will be a third-order interaction mean square whose mathematical expectation more nearly approaches pure measurement error than does the expected mean square for the second-order interaction.

If this unwillingness to *assume* the variance component for the interaction of ratees, raters, and traits inconsequential seems pedantic, note that we are dealing with two sets of individuals, ratees and raters, organisms probably far more likely to interact with each other and with traits than are

many of the variables manipulated by psychologists. While, for example, strong interaction of style of printing type with size of printing type with color of paper may seem quite unlikely, a priori, we cannot in our present state of ignorance about *intra*-individual characteristics afford to assume that second-order interactions involving individuals are infinitesimal.

Statistical Adjustments for Biases of Raters

Guilford ([8], pp. 280–288) recommends that ratings be adjusted to remove the biases due to raters, reflected in significant MS_r , $MS_{(i \times r)}$, and $MS_{(r \times t)}$. His procedure is equivalent to the following, where X'_{irt} represents the adjusted rating of the ith ratee by the rth rater on the tth trait, and $\bar{X}$'s denote means:

$$(29) \quad X'_{irt} = X_{irt} - (\bar{X}_{.r.} - \bar{X}_{...}) - (\bar{X}_{ir.} - \bar{X}_{i..} - \bar{X}_{.r.} + \bar{X}_{...})$$
$$- (\bar{X}_{.rt} - \bar{X}_{.r.} - \bar{X}_{..t} + \bar{X}_{...}).$$

The application of (29) results in adjusted ratings for which MS_r , $MS_{(i \times r)}$, and $MS_{(r \times t)}$ all are zero, but it does not affect $MS_{(i \times r \times t)}$ or the other mean squares. Referring back to (26) and (27), $C - D$ and $F - H$ then become negative: $-MS_{(i \times r \times t)}/T$ and $-MS_{(i \times r \times t)}/I$, respectively. Therefore, Guilford's procedure over-corrects, causing negative bias. The mean covariance across traits within raters becomes less than the mean covariance across traits across raters, representing negative relative halo of magnitude $-MS_{(i \times r \times t)}/T$ when $MS_{(i \times r \times t)}$ is the appropriate error term for $MS_{(i \times r)}$. Similarly, the mean covariance across individuals within traits is made smaller than the mean covariance across individuals across traits.

In order not to over-adjust ratings, one needs a procedure that makes $MS_{(i \times r)}$, $MS_{(r \times t)}$, and MS_r exactly equal to $MS_{(i \times r \times t)}$ without disturbing mean squares other than the three being reduced. This can be done by multiplying each of the two interaction residuals of (29) by the coefficient (1 minus the square root of the ratio of the three-factor interaction mean square to the mean square for the pertinent two-factor interaction):
$1 - \sqrt{MS_{(i \times r \times t)}/MS_{(i \times r)}}$ for the first residual and $1 - \sqrt{MS_{(i \times r \times t)}/MS_{(r \times t)}}$ for the second. Also, for the fixed-effects case, multiply $(\bar{X}_{.r.} - \bar{X}_{...})$ by $1 - \sqrt{MS_{(i \times r \times t)}/MS_r}$. Calling these coefficients a, b, and c, respectively, and simplifying, one obtains a formula that makes the nature of the adjusted scores, X''_{irt} , somewhat clearer:

$$(30) \quad X'_{irt} = X_{irt} + a(\bar{X}_{i..} - \bar{X}_{ir.}) + b(\bar{X}_{..t} - \bar{X}_{.rt})$$
$$+ (a + b - c)(\bar{X}_{.r.} - \bar{X}_{...}).$$

It is easy to show that, by reducing $MS_{(i \times r)}$ to zero, (29) guarantees perfect correlation among raters for total scores of individuals (summed across traits within raters). One estimates the mean correlation among

raters with respect to the sums (or means), over traits, of ratees by ([15], eq. 1)

$$(31) \qquad \bar{r}_{x_{ir}, x_{ir'}} = \frac{MS_i - MS_{(i \times r)}}{MS_i + (R - 1) MS_{(i \times r)}} ,$$

where $r \neq r'$ and

$$X_{ir.} = \sum^{T} X_{irt} \quad \text{and} \quad X_{ir'.} = \sum^{T} X_{ir't} .$$

For $MS_{(i \times r)} = 0$, the right side of (31) reduces to MS_i / MS_i, or unity. Of course the mean r can be unity only when every r between raters is unity. Formula (30) adjusts ratings so as to make $MS_{(i \times r)}$ equal the originally smaller $MS_{(i \times r \times t)}$, thereby increasing the average agreement among raters but not rendering it perfect.

Two scores for each ratee are unaffected by the adjustments of formulas (29) and (30):

$$\sum^{R} X_{irt} \quad \text{and} \quad \sum^{R} \sum^{T} X_{irt} .$$

Therefore, the adjusted trait *sums* (over raters) and adjusted *total* scores (over both raters and traits) cannot be better for any purpose—predictive or otherwise—than the unadjusted ratings were. Furthermore, although the value of $B - D$ in (25) remains constant, both B and D increase equally, while the C of (26) becomes much smaller. In a sense, then, we remove relative halo effect, only to assign it to the general halo effect common to raters without regard to traits.

In fact, the adjustments of (30) typically cause the intercorrelation of the RT rater-trait columns to rise, thereby producing a higher coefficient of equivalence [5] for total scores of individuals across both raters and traits, even though these total scores are not affected at all by the adjustments! This seemingly anomalous result comes about because the adjustment of the $MS_{(i \times r)}$ downward to the magnitude of $MS_{(i \times r \times t)}$ increases the numerator of the following formula for the Hoyt-Cronbach [10, 11, 5] coefficient of equivalence, α, without changing the denominator:

$$(32) \quad \alpha_{x_{i..}} = \frac{MS_i - MS_{(i \times r t)}}{MS_i} = \frac{MS_i - \dfrac{SS_{(i \times r)} + SS_{(i \times t)} + SS_{(i \times r \times t)}}{(I - 1)(RT - 1)}}{MS_i} ,$$

where

$$X_{i..} = \sum^{R} \sum^{T} X_{irt}$$

and where the SS's are sums of squared deviations (i.e., mean squares multiplied by their respective degrees of freedom). Thus the statistical adjust-

ment for relative halo affect cannot affect test-retest or comparable-forms reliability, though when positive halo exists, it does increase the *estimated* internal consistency. (To understand this formula better, see (34) in the Appendix.)

The above paradox arises because one is dealing with a test of the sort that Cronbach [5] calls "lumpy," and also because one treats the new $MS_{(i \times r t)}$ as if it still had $(I - 1)(RT - 1)$ degrees of freedom, when in reality it now has only $(I - 1)(RT - R)$ d.f., because by setting $MS_{(i \times r)}$ at a fixed value— that of $MS_{(i \times r \times t)}$—one loses $(I - 1)(R - 1)$ d.f. The reduction in d.f. may or may not compensate for the reduction in the magnitude of $MS_{(i \times r)}$, so the alpha of (32) might change in either direction. Usually its magnitude will increase.

Though one may have uses for ratings adjusted by (30), such statistical manipulations should by no means substitute for careful designing of the rating study to minimize bias and maximize independence of traits experimentally. Typically, experimental control is superior to statistical control. Where the latter is needed also, Abelson's procedure [1] for finding factors in the traits that maximize $MS_i/MS_{(i \times r)}$ may, when there is significant interaction of ratees with traits, be preferable to (30).

If one had better estimates of error than $MS_{(i \times r \times t)}$, he should use them, instead, for obtaining the a, b, and c that appear in (30). When significant second-order $(i \times r \times t)$ interaction occurs, (30) may adjust too little, this depending upon the appropriate analysis-of-variance model. If each rater rated each ratee $S > 1$ times on each trait, one might employ $MS_{(i \times r \times s \times t)}$, rather than $MS_{(i \times r \times t)}$, for securing a, b, and c, again depending upon the relevant model.

A Numerical Example

Consider Guilford's individuals-raters-traits data ([8], pp. 282–288) from the above point of view. There were 105 ratings, with $I = 7$, $R = 3$, and $T = 5$. Table 1 contains the various mean squares and tests of significance. All main effects and interactions except $MS_{(r \times t)}$ are significantly larger than $MS_{(i \times r \times t)}$ beyond the .05 level.

Applying formulas (5) through (8), $A = 3.351$; $B = 0.763$, $B/A = .23$; $C = 1.851$, $C/A = .55$; and $D = 0.443$, $D/A = .13$. The .23 is identical with the comparable item in Guilford's Table 11.6, and the .55 is almost identical with the mean of the .70, .25, and .74 in the last column of his Table 11.7.

From (2), $MS_{(i \times r)} = (A - B - C + D) + T(C - D) = 8.22$, highly significant when compared with $MS_{(i \times r \times t)} = A - B - C + D = 1.18$ because of the large covariance among traits within raters (C) compared with the small covariance across both raters and traits (D). The mean of the 30 intra-rater coefficients of correlations among traits, estimated by C/A, was

TABLE 1

Analysis of Variance of Ratings of Seven Ratees by Three
Raters on Five Traits, after Guilford ([8], p. 283)*

Source of Variation	d.f.	Mean Square	MS/1.18	P
Among ratees (i)	6	15.82	13.41	< .001
Among raters (r)	2	4.52	3.83	< .05
Among traits (t)	4	11.63	9.86	< .001
$i \times r$	12	8.22	6.96	< .001
$i \times t$	24	2.14	1.81	< .05
$r \times t$	8	1.62	1.37	> .05
$i \times r \times t$	48	1.18	—	—
Total	104	3.56	—	—

*But, using a different procedure for testing significance, Guilford failed to find r or $(i \times t)$ significant.

.55, contrasted with a D/A of only .13. Clearly, strong relative halo effect occurred in this study.

Similarly but less markedly, $\text{MS}_{(i \times t)} = (A - B - C + D) + R(B - D) = 2.14$, significant at the .05 level. The average of the 15 intercorrelations among raters within traits was estimated by B/A to be .23, contrasted with the base-line $\bar{r}$ of .13. Therefore, the traits are to some extent different, though probably not as much as the investigators desired. Finally, $\text{MS}_{(r \times t)}$ is not significant; from formulas (13), (14), and (16) one can estimate, via F/E, that the mean correlation across ratees within traits is $-.03$, contrasted with $-.06$ for the $\bar{r}$ across both ratees and traits, estimated by H/E.

Reducing $\text{MS}_{(i \times r)}$ to the magnitude of $\text{MS}_{(i \times r \times t)}$ via the adjustment in (30) changes B/A from .23 to .51, C/A from .55 to .38, and D/A from .13 to .38. The apparent gain in trait independence is spurious, of course, because both $\text{MS}_{(i \times t)}$ and $\text{MS}_{(i \times r \times t)}$ are unaltered; the $\sum^R X_{irt}$'s are unaffected by the adjustments. Relative halo effect did disappear, being absorbed into the base-line correlation across both raters and traits, reflected by the considerable rise in D/A.

The average of the three r's among raters, estimated by means of (31), changes from .24 for the original rating sums, $\sum^T X_{irt}$, to .81 among such sums of ratings adjusted by (30). The coefficient of equivalence rises from .84 for unadjusted ratings to .89 or .91 for adjusted ones, depending upon how many degrees of freedom, $(I - 1)(RT - R)$ or $(I - 1)(RT - 1)$, are used in (32).

An Extension

For many analysis-of-variance situations one needs a mean square whose mathematical expectation is just σ^2, or very nearly so, in order to devise proper error terms and to estimate components of variance. Having each rater rate each ratee-trait combination more than once under randomized conditions that minimize memory carryover will help meet this need. The multiple ratings of each ratee on each trait can be considered an ordered fourth (fixed?) effect, say sequence, with $s = 1, 2, \cdots, S; S > 1$. Now the notation for the rating received by the ith ratee from the rth rater the sth time on the tth trait is X_{irst}. If the $\text{MS}_{(i \times r \times s \times t)}$ has a relatively large number of degrees of freedom, it might be employed as the MS with $E[\text{MS}] = \sigma^2$, under the reasonable assumption that $\sigma^2_{(i \times r \times s \times t)}$, the component of variance attributable to the third-order interaction, is negligible.

A complete analysis of such ratings, both by analysis-of-variance and correlational methods, may be worthwhile, especially for such comparisons as $\bar{r}_{rst,rs't}$ with $\bar{r}_{rst,r'st}$ to check upon intra-rater versus inter-rater reliability. Components of variance should also be informative. If $S > 2$, one might employ orthogonal polynomials to test for nonlinear trends in the rating sequence [7].

For the four-factor design there are seven mean covariances, as contrasted with three for the three-factor design; these are

$$\overline{\text{cov}\,(X_{rst}, X_{r'st})}, \quad \overline{\text{cov}\,(X_{rst}, X_{rs't})}, \quad \cdots, \quad \overline{\text{cov}\,(X_{rst}, X_{r's't'})}.$$

Because eight mean squares involve ratees, the seven mean covariances and $\bar{s}^2_{rst}$ can be computed.

Concluding Remark

It seems quite likely that the formulas given here are applicable far beyond the ratee-rater-trait situation. Abelson's heuristic table [1] classifying agents, objects, and modes for six types of studies lists the following possibilities from sociometry, clinical ratings, the semantic differential, laboratory experiments, psychological testing, and psychophysical or preference ratings: judges-judgees-items, raters-concepts-scales, conditions-subjects-responses, subjects-conditions-responses (trials?), occasions-subjects-tests, and judges-stimuli-(hypothetical) scale components.

Perhaps approaching a three-way classification of real numbers in the ways suggested in this paper furthers Abelson's goal of offering "a promising combination of experimental and correlational approaches" and partially resolves the dilemma to which Cronbach [6] pointed.

Appendix: Outline of Proof

Gulliksen ([9], p. 54) and Stanley ([15], pp. 90–91) have shown that the $\text{MS}_{(i \times j)}$ of a two-way classification is equivalent to $\bar{s}^2_i - \overline{\text{cov}\,(X_i, X_{i'})}$

where $j \neq j'$. Applying this relationship to the matrix of individuals-by-trait means (over raters), one can by the following procedure secure formula (3):

$$\mathrm{MS}_{(i \times t)} = \sum^I \sum^R \sum^T (\bar{X}_{i \cdot t} - \bar{X}_{i \cdot \cdot} - \bar{X}_{\cdot \cdot t} + \bar{X}_{\cdot \cdot \cdot})^2/(I-1)(T-1)$$

$$= R \sum^I \sum^T \left[\bar{X}_{i \cdot t} - \frac{\sum^I \bar{X}_{i \cdot t}}{I} - \frac{\sum^T \bar{X}_{i \cdot t}}{T} + \frac{\sum^I \sum^T \bar{X}_{i \cdot t}}{IT} \right]^2 /(I-1)(T-1)$$

$$= R \left[\frac{\sum^T s^2_{(\sum^R X_{irt}/R)}}{T} - \frac{\sum^T \sum^{T-1} \mathrm{cov} (\sum^R X_{irt}/R, \sum^R X_{irt'}/R)}{T(T-1)} \right]$$

$$= \left[\sum^T s^2_{(X_{1t} + X_{2t} + \cdots + X_{Rt})} \right.$$

$$\left. - \frac{\sum^T \sum^{T-1} \mathrm{cov} (X_{1t} + \cdots + X_{Rt}, X_{1t'} + \cdots + X_{Rt'})}{T-1} \right] /RT$$

$$= \left\{ \sum^T \left[\sum^R s^2_{rt} + \sum^R \sum^{R-1} \mathrm{cov} (X_{rt}, X_{r't}) \right] \right.$$

$$\left. - \sum^T \sum^{T-1} \left[\sum^R \mathrm{cov} (X_{rt}, X_{rt'}) + \sum^R \sum^{R-1} \mathrm{cov} (X_{rt}, X_{r't'}) \right]/(T-1) \right\} /RT$$

$$= \overline{s^2_{rt}} + (R-1) \overline{\mathrm{cov} (X_{rt}, X_{r't})} - \overline{\mathrm{cov} (X_{rt}, X_{rt'})}$$

$$- (R-1) \overline{\mathrm{cov} (X_{rt}, X_{r't'})} = A + (R-1)B - C - (R-1)D.$$

Formulas for the other first-order interactions can be obtained in the same way as (3), above.

To secure (4), for $\mathrm{MS}_{(i \times r \times t)}$,

(33) $\mathrm{MS}_{(i \times r t)} = A - [(R-1)B + (T-1)C$

$$+ (R-1)(T-1)D]/(RT-1)$$

and then

(34)   (Sum of Squares)$_{(i \times r t)} = \mathrm{SS}_{(i \times r)} + \mathrm{SS}_{(i \times t)} + \mathrm{SS}_{(i \times r \times t)}$.

Formulas (1), (9), and (17) are readily secured in a straightforward manner from the definitional formulas for MS_i, MS_r, and MS_t. Finally, note that, for example,

(35) $RT(I-1)A = \mathrm{SS}_i + \mathrm{SS}_{(i \times r)} + \mathrm{SS}_{(i \times t)} + \mathrm{SS}_{(i \times r \times t)}$.

This relationship, known from fundamental considerations of the analysis of variance *before* solving for A via formulas (1)–(4), constitutes an independent check of (5) and, therefore, indirectly of (6)–(8).
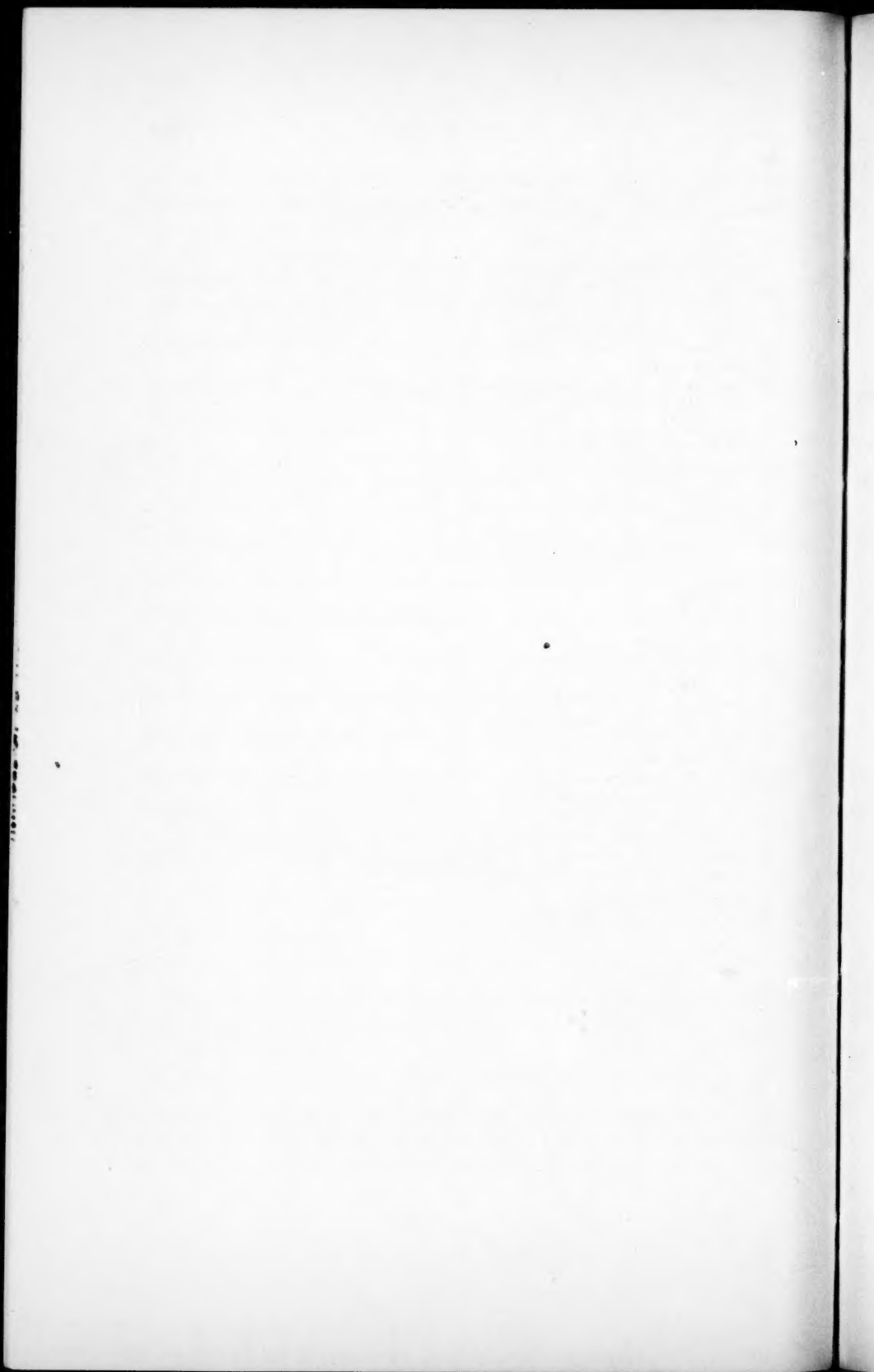
REFERENCES

[1] Abelson, R. P. A discriminant approach to factoring three-way data tables. *Amer. Psychologist*, 1958, **13**, 375. (Abstract) (More extensive reports privately circulated.)

[2] Campbell, D. T. and Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, **56**, 81–105.

[3] Chi, P.-L. Statistical analysis of personality rating. *J. exp. Educ.*, 1937, **5**, 229–245.

[4] Cornfield, J. and Tukey, J. W. Average values of mean squares in factorials. *Ann. math. Statist.*, 1956, **27**, 907–949.

[5] Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, **16**, 297–334.

[6] Cronbach, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, **12**, 671–684.

[7] Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

[8] Guilford, J. P. *Psychometric methods* (2nd ed.) New York: McGraw-Hill, 1954.

[9] Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.

[10] Hoyt, C. J. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **6**, 153–160.

[11] Hoyt, C. J. and Stunkard, C. L. Estimation of test reliability for unrestricted item scoring methods. *Educ. psychol. Measmt*, 1951, **12**, 756–758.

[12] Humphreys, L. G. Note on the multitrait-multimethod matrix. *Psychol. Bull.*, 1960, **57**, 86–88.

[13] Johnson, D. M. and Vidulich, R. N. Experimental manipulation of the halo effect. *J. appl. Psychol.*, 1956, **40**, 130–134.

[14] Stanley, J. C. Fixed, random, and mixed models in the analysis of variance as special cases of a finite model. *Psychol. Rep.*, 1956, **2**, 369.

[15] Stanley, J. C. K-R 20 as the stepped-up mean item intercorrelation. *14th Yrbk Natl Coun. Meas. used in Educ.*, 1957. Pp. 78–92.

[16] Wilk, M. B. and Kempthorne, O. Fixed, mixed, and random models. *J. Amer. statist. Ass.*, 1955, **50**, 1144–1167.

[17] Wilk, M. B. and Kempthorne, O. Derived linear models and their use in the analysis of randomized experiments. *WADC Tech. Rep.* 55-244, Vol. II, Mar. 1956.

[18] Wilk, M. B. and Kempthorne, O. Some aspects of the analysis of factorial experiments in a completely randomized design. *Ann. math. Statist.*, 1956, **27**, 950–985.

[19] Willingham, W. W. and Jones, M. B. On the identification of halo through analysis of variance. *Educ. psychol. Measmt*, 1958, **18**, 403–407.

MULTIDIMENSIONAL UNFOLDING: DETERMINING CONFIGURATION FROM COMPLETE RANK ORDER PREFERENCE DATA

WILLIAM L. HAYS

UNIVERSITY OF MICHIGAN

AND

JOSEPH F. BENNETT*

LINCOLN LABORATORY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Within the model of isotonic space, a principle is presented which generalizes the unfolding technique to the multidimensional case. The availability of exhaustive configurational solutions given complete data is pointed out. Finally three criteria are suggested for the choice of a particular solution from among the set of all solutions, which are applicable in the case either of complete or incomplete data.

In an earlier paper [2], the authors discussed the estimation of the dimensionality underlying a set of rank orders. The data to which these methods are applicable are rank orders of preference given by subjects for a set of n objects. The dimensionality estimated is that required by the stimulus space in which the objects are presumably viewed by the judges. Three lower-bound criteria for dimensionality were proposed: mutual boundary, cardinality, and the existence of permutation groups. The present paper continues with the generalization of Coombs' unfolding technique [3] to the multidimensional case, which was begun in the first paper, and concerns the problem of determining configuration and arriving at a solution when the data are "complete," in the sense to be described below. In particular, a principle providing the generalization of the unfolding method will be introduced, and various criteria for determining a set of r rank order axes for the description of a configuration of n stimulus points in r dimensions will be discussed. It will be convenient to acquaint the reader with some terminology before the matter of a solution is taken up.

Some Terminology

This discussion will be couched in terms of a simple generalization to several dimensions of the model proposed by Coombs [3, 4]. It is assumed

*Deceased.

that subject i views stimulus k as a point X_{ik} in a stimulus space of some dimensionality r. Given some arbitrary origin, the components of X_{ik} are x_{ijk} , the projection or loading of stimulus k for individual i on attribute j. To simplify matters, it will be assumed that x_{ijk} is the same for each individual i, so that the projection of stimulus k on attribute j may be denoted simply by x_{jk} , a component of X_{ik} . The stimulus space will be assumed to have Euclidean properties as well. Strictly speaking, neither of these assumptions may be essential, but it has proved very difficult to develop the model without some such restrictions.

Also following Coombs, it is assumed that each individual i may be associated with an "ideal" stimulus, a real or hypothetical object which the individual would most prefer in any given stimulus space. This ideal stimulus may also be represented as a vector C_i with components c_{ij} . Once again the simplifying assumption is made that each individual is associated with one and only one such ideal point in the space.

Now the observational equation linking the judged rank order of preference to the distances within the stimulus space is given by

$$(k \cdot \succ m)_i \Leftrightarrow \sum_{j=1}^{r} (c_{ij} - x_{im})^2 > \sum_{j=1}^{r} (c_{ij} - x_{ik})^2.$$

That is, stimulus k will be preferred to stimulus m by individual i if and only if the sum of the squared differences between a stimulus and the ideal stimulus over a set of r orthogonal reference axes is greater for stimulus m than for stimulus k. The subject prefers that stimulus which is closer to his ideal in the stimulus space.

In short, the stimuli are conceived simply as having some configuration in a Euclidean stimulus space of dimensionality r. A particular rank order of preference reflects increasing magnitudes of distance from the ideal stimulus to the respective stimulus points. Obviously, for other than the one-dimensional case, a solution to this problem of describing the stimulus configuration requires an excursion into the geometry of higher spaces. Furthermore, since the data with which we start are nothing more than a set of rank orders, a special class of such higher spaces must be considered; these are so-called *isotonic* spaces, in which every region in the space is characterized by a rank order of distances to a fixed set of points. In other words, given the set of stimulus points, each and every point in the space must show a rank order of distances to these stimulus points, and the space as a whole may be divided into isotonic regions, convex subspaces within which each and every point shows the same rank order of distances from the stimulus points. The rank order which each point in the region exhibits will be called the characteristic order for the region, and any region will be referred to by its characteristic order.

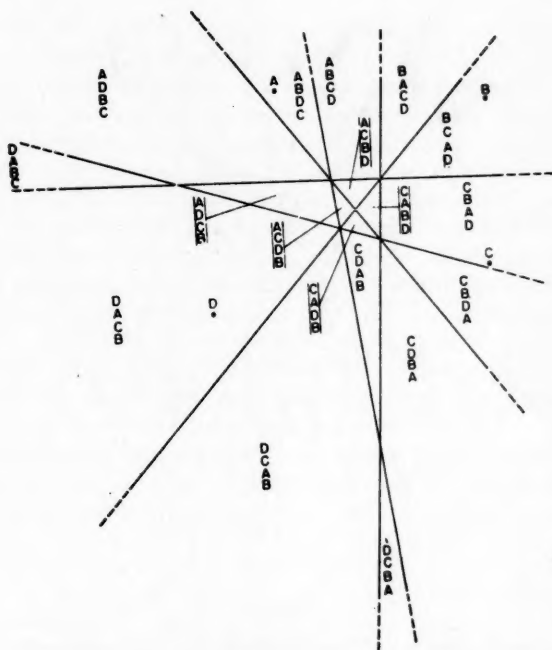The one-dimensional case of an isotonic space is, of course, the under-

FIGURE 1

A Configuration of Four Points in Two Dimensions

lying model for the unfolding technique of Coombs [3]. The unfolding technique is based upon the principle that a sequence of $\binom{n}{2} + 1$ rank orders may be found representing the unique arrangement of the isotonic regions in a one-dimensional space; from this sequence of rank orders, the order of stimulus points on the attribute may be inferred, as well as a partial order of the distances between points.

There is no such unique sequence in the case of two or more dimensions, however, as the number of possible isotonic regions increases very rapidly both with the number of stimulus points and the number of dimensions. An example of an isotonic space for four points in two dimensions is given by Figure 1, and an example for five points also in two dimensions by Figure 2.

The three criteria of dimensionality adverted to above may be illustrated from these examples. Note, for instance, that no more than four isotonic regions anywhere mutually bound in either Figures 1 or 2: this reflects two dimensions, according to the first criterion. Second, notice that only 18 of the 24 possible permutations of four objects occur as charac-
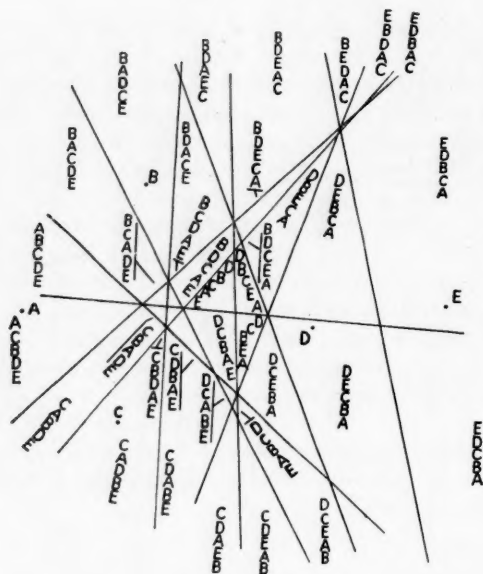
FIGURE 2

Isotonic Regions Generated by Five Points in Two Dimensions

teristic orders for regions in Figure 1, and that only 44 different orders from among the 120 possible permutations of five things occur in Figure 2: this illustrates the so-called *cardinality* criterion. Finally, and most important, note that while there are complete sets of 6 permutations in rank order for subgroups of three stimuli embedded among the rank orders, there exist no such complete sets for subsets of either four or five stimuli: this illustrates the *groups* criterion, which will be of importance in determining configuration as well as in estimating dimensionality for such data.

The isotonic regions are bounded by loci of equidistance from two stimuli—these represent regions of equal preference for the two stimuli. The loci of equidistance in two-dimensional cases as illustrated are merely the perpendicular bisectors of the lines joining pairs of stimulus points. These appear in the figures as the lines dividing the space into regions. Such loci of equidistance from pairs of stimuli will be referred to as 2-loci, and will be denoted by $H(A, B)$, where A and B refer to a particular pair of stimulus points. For any configuration there will be $\binom{n}{2}$ such 2-loci. It is essential to remember that in one dimension, a 2-locus will be a point, in two dimensions a line, in three dimensions a plane, in four dimensions a.

three-space, and so on. In general, in r dimensions, the 2-locus $H(A, B)$ will be a hyperplane of dimensionality $r - 1$.

Another feature of note in any isotonic space is the fact that regions may be divided into two disjoint classes: *open* and *closed*. Closed regions are, of course, everywhere bounded by loci of equidistance, while open regions are not. The distinguishing feature of rank orders derived from open and closed regions is that *each and every open region must have a mirror image mate*, another region which has a characteristic order which is the exact reverse of the first rank order; any pair of mirror image rank orders in the data leads to the inference of the existence of a pair of open regions. On the other hand, *no closed region may have such a mirror image mate*. In any dimensionality lower than $n - 1$ for n stimuli, there must exist some closed regions in the isotonic space, and hence all $n!$ rank order permutations may not occur in the data for less than dimensionality $n - 1$. This is actually the basis for the cardinality criterion of dimensionality.

It will be noted from Figures 1 and 2, that there are points of intersection of sets of three 2-loci. These intersections are equidistant from sets of three stimuli, and as such are appropriately called 3-loci and designated $H(A, B, C)$, where (A, B, C) is any set of three stimuli. In two dimensions a 3-locus is a point, in three dimensions a line, in four dimensions a plane, and so on. In general, in r dimensions, a 3-locus will be a hyperplane of $r - 2$ dimensions.

It will be convenient to consider loci of even higher order, so that a general notation and dimensional principle will be useful. A g-locus will be the space of all points equidistant from a set of g points, $H(A, B, C, \cdots, g)$, and will always be a subspace of dimensionality $r + 1 - g$. Furthermore, it will also be useful to remember that a point which is equidistant from some g stimuli in $g - 1$ dimensions is the center of a hypersphere having the stimulus points in question on its surface. Thus, three points requiring two dimensions must lie on a circle, four points requiring three dimensions must lie on a sphere, five points requiring four dimensions must lie on a four-dimensional hypersphere, and so on. Furthermore, the converse is true— if there exists no hypersphere in $g - 1$ dimensions such that a particular set of g points may lie on its surface, then the set of points may be embedded in a space of $g - 2$ dimensions or less. As will appear in the discussion to follow, this is simply another way of phrasing the groups criterion of [2].

One final feature of the higher spaces should also be mentioned here: given a subspace T of dimensionality t and a subspace S of dimensionality s, $t \geq s$, such that S is not a subspace of T, then the intersection of S and T is a subspace of dimensionality $s - 1$. Thus, the intersection of a plane and a line (not entirely in the plane) is a point, the intersection of two non-coincident planes is a line, the intersection of a six-space and a three-space is a plane, and so on.

A General Unfolding Principle

As mentioned above, the one-dimensional unfolding technique relies on the fact that if the rank orders emanate from a one-dimensional stimulus space, it is always possible to construct a unique sequence such that each distinct rank order differs from either neighbor in the sequence by a reversal in order of only one pair of objects. The end or mirror image rankings then provide the order of the objects on the attribute.

Coombs' procedure could perfectly well be interpreted as finding a sequence of 2-loci rather than a sequence of isotonic regions, since the reversal in order of a pair of objects for a pair of regions simply fixes such a 2-locus point. Considering the unfolding technique in this way suggests the principle which allows an extension to the multidimensional case. Since this principle is in fact the general statement of Coombs' basic idea, it seems important and nontrivial enough to state and prove.

PRINCIPLE. Given some fixed line L, and three points A, B, and C in general position in an isotonic space of dimensionality r, the line L intersects the 2-loci $H(A, B)$, $H(B, C)$, and $H(A, C)$ in that order (or the reverse) if and only if the perpendicular projections of the three points upon L are in the order CAB or the reverse.

PROOF. The necessary condition will be proved first. If the line L coincides with the first of a set of r orthogonal reference axes $(X_1, X_2, \cdots, X_r)$, such that the origin lies at the intersection of L and $H(A, B)$, then every point k on L is characterized by an r-tuple $(x_{1k}, 0, \cdots, 0)$. Let the intersection of L and $H(A, B)$ be $(x_{11}, 0, \cdots, 0)$, with $x_{11} = 0$, that of L and $H(B, C)$ be $(x_{12}, 0, \cdots, 0)$, and L and $H(A, C)$ be $(x_{13}, 0, \cdots, 0)$. Finally let the point A be characterized by $(x_{1A}, x_{2A} \cdots, x_{rA})$, and similarly for B and C. Now the 2-locus $H(A, B)$ is defined by

$$(1) \quad (x_1 - x_{1A})^2 + \cdots + (x_r - x_{rA})^2 = (x_1 - x_{1B})^2 + \cdots + (x_r - x_{rB})^2,$$

where $(x_1, x_2, \cdots, x_r)$ is any point lying in the 2-locus. Similar definitions may be made for the other two 2-loci as well. Solving (1) for x_{11} and putting $(x_1, x_2, \cdots, x_r) = (x_{11}, 0, 0, \cdots, 0)$ gives

$$(2) \quad x_{1A}^2 + \cdots + x_{rA}^2 = x_{1B}^2 + \cdots + x_{rB}^2.$$

The value of x_{12} is given by

$$(3) \quad x_{12} = \frac{x_{1B}^2 - x_{1C}^2 + \cdots + x_{rB}^2 - x_{rC}^2}{2(x_{1B} - x_{1C})}$$

and the value of x_{13} by

$$(4) \quad x_{13} = \frac{x_{1A}^2 - x_{1C}^2 + \cdots + x_{rA}^2 - x_{rC}^2}{2(x_{1A} - x_{1C})}.$$

Because the intersections of the 2-loci with L are in the order $H(A, B)$,

$H(B, C)$, $H(A, C)$ or the reverse, the absolute difference between x_{11} and x_{12} must be less than that between x_{11} and x_{13} , or $d_{13}^2 > d_{12}^2$, where

$$(5) \qquad d_{13}^2 = \frac{(x_{1A}^2 - x_{1C}^2 + \cdots + x_{rA}^2 - x_{rC}^2)^2}{4(x_{1A} - x_{1C})^2} ,$$

$$(6) \qquad d_{12}^2 = \frac{(x_{1B}^2 - x_{1C}^2 + \cdots + x_{rB}^2 - x_{rC}^2)^2}{4(x_{1B} - x_{1C})^2} .$$

However, subtracting $x_{1C}^2 + x_{2C}^2 + \cdots + x_{rC}^2$ from each side of (2) and squaring shows that the numerators of (5) and (6) must be equal. Hence it follows that since d_{13}^2 exceeds d_{12}^2 , the denominator of (6) must exceed the denominator of (5), so that

$$(x_{1B} - x_{1C})^2 > (x_{1A} - x_{1C})^2,$$

the perpendicular projections of A and C upon L must be nearer than those of B and C. An identical argument using $x_{13} = 0$ shows that the distance between A and B projections must also be less than that between projections of B and C. Since distance is invariant under translation or rotation of axes, the necessary condition is proved. The sufficient condition is proved simply by reversing the steps of the necessary condition argument.

The unfolding technique for one dimension is actually a special case of this more general principle for finding projections upon lines by constructing sequences of regions (or dually, 2-loci). In isotonic space of any dimensionality, the existence of $\binom{n}{2} + 1$ regions which fit the unfolding qualifications is sufficient for the inference of the order of projections which the points have relative to some line. As an illustration of this principle, consider the following sequence of seven regions drawn from the example of Figure 1: $DACB$, $DCAB$, $CDAB$, $CDBA$, $CBDA$, $CBAD$, $BCAD$. In this sequence, $H(C, D)$ lies between $DCAB$ and $CDAB$, $H(A, C)$ falls between $DACB$ and $DCAB$, and $H(A, D)$ lies between $CBDA$ and $CBAD$; the order of these three 2-loci is thus $H(A, C)$ $H(C, D)$ $H(A, D)$ (or the reverse), so that on a line extending through these seven regions, the order of projections of the three stimulus points A, C, and D must be CAD (or the reverse). Likewise, the order of A, B, and C as projected upon such a line would be ACB or the reverse, since the order of their 2-loci is $H(A, C)$ $H(A, B)$ $H(B, C)$. An inspection of the 2-loci for all such triples of stimuli establishes that the order of projections on such a line would be $DACB$ or the reverse (obviously, since there is no fixed origin in the isotonic space, the orders of projections on any line may be read in either direction). Any line capable of being located in the space must pass through such a sequence of $\binom{n}{2} + 1$ regions, and any complete unfolding sequence of $\binom{n}{2} + 1$ regions occurring in the space must represent at least one possible line in the space.

An important feature of Coombs' unidimensional unfolding solution is the recovery of metric relations among the stimulus points, yielding an ordered metric scale of the stimuli. This metric information is inferred from the sequence of 2-loci just as is the simple order of the stimuli themselves. Unfortunately, it can be shown that the ability to obtain metric information in this way is restricted to the unidimensional case. While the theorem above is a complete generalization of the method for obtaining the order (or the order of projections) of the stimulus points, in other than the one-dimensional case, Coombs' method for inferring metric information does not work for the *projections* of the points upon axes in the space. An important corollary follows directly from the principle just given.

> In any isotonic space of n stimuli in r dimensions, two regions may have characteristic orders which are mirror images if and only if there exists the possibility of a line in the space such that the order of projections of the stimuli on the line is the same as the characteristic order of either of the regions (or the reverse, of course).

An Exhaustive Solution for Complete Data

The practical implication of this corollary principle is that any pair of mirror images existing in the data afford a potential solution, in that there must exist the possibility of an axis showing such an order of projections. Even more important is the fact that any potential solution must be represented by such a mirror image pair of regions in the data, when the data are complete. In this light, the question of a solution for complete data becomes rather trival. First, in the present context, let *complete data* be understood to mean sets of rank orders such that each and every isotonic region in the stimulus space has its characteristic order represented at least once in the data. Thus when complete data are at hand, all possible configurational solutions may be recovered from the data simply by finding mirror image pairs of rankings. Each mirror image pair located is one potential axis for describing the configuration; for this reason the solution from complete data may be called *exhaustive*.

In the simple example of Figure 1, the mirror image pairs of regions are *DACB–BCAD, DABC–CBAD, ADBC–CBDA, ABDC–CDBA, ABCD–DCBA, BACD–DCAB*. Thus there are six possible simple orders which may represent axes or solutions to the configuration: *DACB, DABC, ADBC, ABDC, ABCD,* and *BACD* (or their reverses). These constitute the exhaustive solution for this configuration.

The number of such potential axes varies, of course, both with the number of stimuli and the dimensionality. Actually, it is possible to calculate the maximum number of distinct such axes (i.e., distinct mirror image pairs

of regions). These maximum numbers have already been tabled in another context [1] in the form of the maximum number of open isotonic regions which may exist for given numbers of stimuli and dimensionalities; in order to convert this table into maximum number of rank order axes, one simply divides the entries by 2. Thus, for example, there are 36 different rank orders of projections possible for 5 stimuli in three dimensions, 105 different possible rank order axes for 15 stimuli in two dimensions, and just over a billion possible rank orders of projections for 30 stimuli in five dimensions! Obviously, such exhaustive solutions leave something to be desired in the way of parsimony of description. Moreover, seldom would we be interested in all solutions anyway, even if there were fairly restricted numbers of such possibilities.

Under the influence of factor-analysis methods, we have grown accustomed to the description of configurations of points requiring r dimensions in terms of a set of axes numbering fewer than r. While the usual factor analysis deals only with common factors, no such restriction exists within this model. The dimensionality estimated for an isotonic space includes both common and specific factors and, in principle, it should be possible to analyze the data for all r dimensions. Still another difference exists between metric and nonmetric approaches to this problem: metric methods such as factor analysis provide dimensions from which one may reproduce the original data, while, at this writing, there seems to be no prospect that one might reproduce an original set of rank orders in terms of some r rank order dimensions. In a sense there is more information in the data than in the rank order dimensions obtained. Failing any criterion for the reproducibility of the data in nonmetric terms, the only recourse seems to be to choose some solution from among the set of all solutions according to criteria of a best fit to the data. In order to do this, one must settle upon some criteria of goodness for choosing among all possible solutions.

The remainder of this paper will be devoted to a description of three features of the isotonic space which may serve as criteria in the choice of a rank order solution from all possible such solutions. These criteria are applicable not only in the theoretical case of complete data but also in the case of incomplete data as well, and thus they will be described in detail here.

The Idea of a Central Intersection

One requirement for a solution might be that each successive axis pass through the *center* or greatest concentration of points in the configuration. That is, the first axis should describe the *length* of the configuration through its greatest concentration; the second axis should describe the length of the configuration of projections of points on a space of one less dimension, and so on. Thus, axes may be sought which are roughly analogous to principal axes in the usual factor-analysis model.

In order to find such a solution within the isotonic space, the idea of a *least intersection* of a configuration may be introduced. In any r-dimensional space, the locus of equidistance from any $r + 1$ points in general position is a point. This point is the center of a hypersphere of dimensionality r. Thus, the locus of equidistance from three points in two dimensions is the center of a circle (2-sphere); the locus of equidistance from four points in three dimensions is the center of a sphere (3-sphere); the locus of equidistance from five points in four dimensions is the center of a hypersphere of four dimensions (4-sphere), and so on.

Any hypersphere of whatever dimensionality must bear one of three possible relationships to any point in the space: the point in question must either be interior to the hypersphere (fall within its surface), exterior to the hypersphere (fall beyond the space enclosed by its surface), or conjoint with the hypersphere (fall upon its surface, and thus be at a distance from its center equal to that of any other point on the surface). Furthermore, if a stimulus point X is exterior to an r-sphere, then the order of distances which is characteristic of its center must show the point X more distant from the center than any point on the surface. On the other hand, if the point X is interior to the r-sphere, then the order associated with the center must show X less distant from the center than any point on the surface. Finally, if the point X is conjoint with the hypersphere, the order associated with the center must show the X equally distant with any point on the surface.

Obviously, among n stimulus points in general position in r dimensions, any given stimulus point must be conjoint with $\binom{n-1}{r}$ distinct r-spheres, since there will exist a center of an r-sphere for each set of $r + 1$ points. Also, any pair of points must be conjoint with $\binom{n-2}{r-1}$ r-spheres.

The various r-spheres defined by the points in the space will differ in the extent to which they include the entire configuration within or on their surface. Some r-spheres will have none of the remaining $n - r - 1$ points interior to its surface. At least one r-sphere will have all of the points either within or on its surface. Such an r-sphere containing all points either within or on its surface will be called an *enveloping sphere*. Given two or more enveloping spheres, the subspace formed by the intersection of the spaces they bound will contain the configuration; this subspace will be called the *central intersection* of the spheres. In Figure 3, for example, the circle defined by the points A, C, and E is an enveloping sphere, since all five points are either on or within its surface, and the same is true of the circle defined by A, B, and E. The central intersection of these two circles, as shown by the shaded area, contains the configuration.

In particular, if there exist r distinct enveloping r-spheres each generated by the same two points X and Y and some set of $r - 1$ other points,
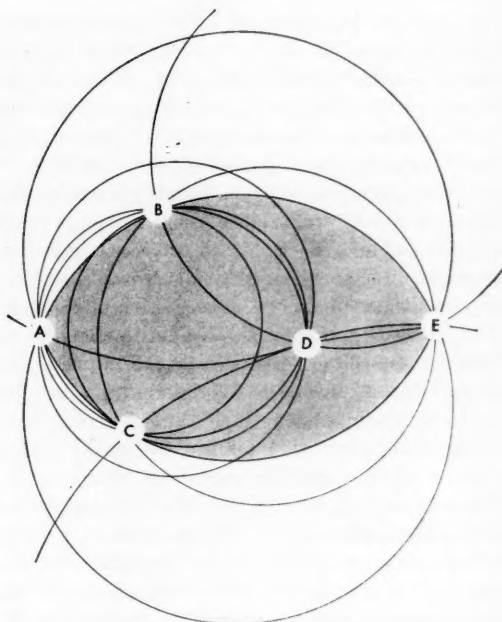
FIGURE 3

Circles Generated by Five Points in Two Dimensions Showing the Least Intersection
for the Configuration

then the central intersection of all of the $\binom{n-2}{r-1}$ spheres will be called a
least intersection. The major axis of this least intersection subspace will be
the line joining X and Y, and the minor axes of the subspace must lie in
the hyperplane of equidistance $H(X, Y)$ defined by these two points. In
Figure 3, the shaded area bounded by the two circles formed by A, B, and
E and by A, C, and E is a least intersection, and the major axis of this area
is the line $\overline{AE}$ with a minor axis defined by their perpendicular bisector.

However, since one is dealing strictly with an isotonic space, in which
the only information available is in the form of rank orders for regions, the
problem remains of finding that pair of points defining the axis of a least
intersection. This may be done as follows. The center of an r-sphere is, of
course, an $(r + 1)$-locus, the point of equidistance from some set of $r + 1$
points. The $(r + 1)$-locus point does not fall into any isotonic region having
a simple characteristic order of distance; rather, such an $(r + 1)$-locus must
fall on the boundary separating a number of such regions, and consequently
have a partial ordering of distances, since it is by definition equally distant

from at least $r + 1$ points. For example, in the configuration of five points in two dimensions, the 3-locus $H(A, B, C)$ for the three points A, B, and C is the point of intersection of the three 2-loci $H(A, B)$, $H(A, C)$, and $H(B, C)$, thus falling on the boundary lines among six regions. Both D and E are exterior to the circle, so that the order of distances from the center of the circle to the five points is the partial order $(ABC)DE$. (Single parentheses enclosing a set of points in a partial order will always denote equality among the points in the set, while double parentheses will denote the set of all permutations in order of the enclosed set: thus $(ABC)DE$ is read as A, B, C equally in the first place followed by D and E, while $((ABC))DE$ would read as any of the set of six orders consisting of some permutation of A, B, and C followed by D and then E.) Note also that the six regions immediately surrounding the center of this circle are all alike in order, except that each shows a different one of the six permutations $((ABC))DE$, the positions of D and E remaining fixed. Such a set of *permuting* regions would also be used in the groups criterion for dimensionality mentioned earlier. The general principle which such sets of regions exhibit is: if there exist a set of $(r + 1)!$ regions showing exactly the same characteristic orders except for a permutation of some set of $r + 1$ stimulus points, then there exists an r-sphere to which those $r + 1$ stimuli are conjoint. The partial order of distances characteristic of the center of the r-sphere is the same as that of any of the set of permuting regions, except that all of the set of conjoint stimuli are equally distant from the center point, thus giving it a partial order of distances. In other words, the groups of permutations occurring in the data tell not only about the dimensionality, they also tell of the existence of r-spheres in the space. Figures 2 and 3 again provide an illustration. Each one of the ten circles is accompanied by a set of six permuting regions, and each set of permuting regions surrounds the center of a circle. No set of 4! permuting regions may be found, however, since 2 is the dimensionality.

Furthermore, these sets of permuting regions also give information about the positions of all of the points relative to each of the circles. It has already been mentioned that the set of $r + 1$ stimuli which permute among the $(r + 1)!$ rank orders are those which are on the r-sphere. If any stimulus falls in order below the permuting stimuli for the region orders, then that stimulus is necessarily exterior to the r-sphere. On the other hand, if any stimulus point precedes the stimuli which permute in order, then that stimulus point is interior to the circle. For instance, the regions $CBADE$, $CBDAE$, $CDBAE$, $CDABE$, $CADBE$, $CABDE$, which are members of the set $C((ABD))E$, differ only by a permutation of A, B, and D; thus they must surround the center of a circle in the example, with A, B, and D on the perimeter. Since among these regions C always precedes A, B, and D, the circle must have C as an interior point. However, E always follows the three permuting stimuli in all of the regions of the set of six, so that one must

conclude that the circle will have E exterior to it. This may be seen from Figures 2 and 3.

Since the positions of the points relative to the r-spheres may be read from the orders characterizing such permutation groups in the data, a way emerges for fixing the least intersection for a configuration of points. Recall that the least intersection is the subspace formed by the intersection of r distinct r-spheres such that each point is either conjoint with or interior to each r sphere. Then a least intersection may be determined by first finding all those permuting sets of $(r + 1)!$ orders which show the property that all stimulus points are either in the permuting set or precede the permuting set in order. That pair of stimulus points common to the permuting set for r such groups of stimuli is the major axis of the least intersection. In the example, these r-spheres are represented by the sets of six regions falling into the partial order $CD((ABE))$ and the set of six regions falling into the partial order $BD((ACE))$. The points common to the permuting set for both groups are A and E; consequently A and E describe the long axis of the least intersection for this configuration.

Moreover, the two points which define the major axis of the least intersection of the space have a property which permits them to be identified simply, without the necessity of inspecting all of the sets of $(r + 1)!$ permutation groups of regions. The circumstance that these two points are conjoint with the r-spheres forming the least intersection makes it true that among the closed regions of the space (i.e., those having no mirror image), this particular pair of points will appear in the last two places in order for the largest number of regions. On the other hand, the desired pair will appear in neither the first two places nor the last two places in any open region. Thus, the endpoints of the major axis of the least intersection may be found very easily for complete data by merely counting the number of times pairs of stimulus points appear in the last places for closed regions, minus the number of times the pair appear in an extreme position (at either end of the order) for open regions.

For example, in Table 1 based on Figure 2, it can be seen that the pair A and E occurs in last place in nine of the closed regions, and in an extreme position in none of the open regions. Thus, the major axis of the least intersection must terminate in A and E.

To recapitulate, one criterion which is proposed for choosing among rank order solutions is that the axes chosen reflect the tendency of the points to cluster along the long axis of the least intersection, so that the axes chosen may reflect the general shape of the configuration in so far as possible. This will be taken as the first requirement for a choice from among all of the available solutions for complete data.

Albeit the axis of the least intersection will be determinate for most configurations, it is possible to construct configurations in which there will

TABLE 1

Open and Closed Regions for the Five-Point Example

Open Regions		Closed Regions	
ABCDE	EDCBA	BDACE	BCADE
ACBDE	EDBCA	BCDAE	BDCAE
BACDE	EDCAB	BDCEA	BDEAC
BADCE	ECDAB	BDECA	BDAEC
BADEC	CEDAB	BEDCA	CBDAE
BAEDC	CDEAB	DCBAE	DCABE
BEADC	CDAEB	CDEBA	CDBEA
BEDAC	CADEB	DCBEA	DBCAE
EBDAC	CADBE	DBCEA	DBECA
EDBAC	CABDE	DEBCA	DECBA
		DCEBA	EBDCA
		ECDBA	CEDBA

be fewer than r distinct r-spheres, each of which will have the property of including all of the points within or on its surface. In this situation, there will be ambiguity as to which of the pairs of points best characterize the major axis of the configuration. For example, in the configuration of Figure 2, if one ignores stimulus A, and concentrates on B, C, D, and E only as four points in two dimensions, he can see that the circle generated by B, C, and E fits the qualification for one enveloping circle, but that there is no circle among the remaining three described by B, C, and D; B, E, and D; and C, E, and D which fits this qualification. For this configuration, then, there is no special choice among the pairs BC, BE, and CE as determining the end points of a first axis. There will always, however, exist at least one enveloping r-sphere in any configuration in any dimensionality r.

A Quasi-Simple Structure Criterion

It does not seem quite enough, however, to insist that the rank order axes chosen should reflect the length, breadth, and height of the configuration. It seems desirable to seek solutions which have a certain degree of inherent parsimony of description. In other words, another aspect to a good solution should be its simplicity in some sense. This is true especially since there seems to be no good analogy to rotation within the isotonic model.

The search for such solutions in factor analysis is indissolubly linked with the name of Thurstone and the concept of simple structure [5]. While the rules for achieving simple structure seem to be very much bound up with the mechanics of factor analysis, and especially of rotation, there does seem to be one aspect which may have a rough analogy in the isotonic model. In describing the characteristics of simple structure, Thurstone ([5], p. 335)

emphasized the desirability of maximizing the number of zero loadings which any given factor should show, while, at the same time, minimizing the number of factors on which a test should show high loadings. In the isotonic model, there is, of course, no unique origin in the space, and since the possible solutions are only ordinal in character, the concept of zero loading has no special meaning for this model. However, it does seem that a pertinent part of this requirement for simple structure is not that the loadings for a number of tests are zero per se, but rather that the differences among a maximal number of points are zero when projected upon an axis. In other words, since reference axes are ways of describing the differences which exist among points anyway, maximum clarity is achieved when the various axes describe different sorts of differences among the points of the configuration, so that differences which project large upon one axis shall not project large upon others. This is emphatically not the only interpretation of the concept of simple structure by any means; it is, however, an aspect for which there is at least a distant analogy within the confines of an isotonic space. Thus, a quasi-simple structure requirement may also be imposed in the choice of a solution: each axis should be chosen in such a way that the number of zero distances among projections on each of the axes is maximal.

For n points in r dimensions, there are, in a sense, $\binom{n}{r}$ ready-made axes consisting of all of the r-loci in the space. Recall that the locus of equidistance from r points in r dimensions is a line, and this line must have projections of all of the points upon it. Hence, each of the r-loci is a potential axis. Furthermore, these r-loci do have one valuable property in the light of the quasi-simple structure notion just introduced. This is that the r points defining the locus must project onto exactly the same point upon it; that is, since they are all equally distant from any point on the r-locus, their projections onto the locus must coincide. For example, note in the five-point example that on the line $H(A, B)$ for instance, how the projections of A and of B must coincide, and so on for each pair of points defining a 2-locus line. Consequently, this quasi-simple structure criterion may be approached by taking the r-loci themselves as the axes. This reduces the choice of the set of r axes to some extent, but there are still $\binom{n}{r}$ such loci from which to choose.

How may one determine the order of projections upon an r-locus from complete data? The answer is, by finding permutation groups in sets of $r!$ open regions. According to the corollary of the theorem, open regions must describe possible orders of projections, and permuting sets of $r!$ open regions must thus describe the order of projections upon the line described by an r-locus. In the example, the line $H(A, B)$ has an order of projections

$(AB)CDE$, which is reflected by the set of open regions $ABCDE$ and $BACDE$, a permuting set $((AB))CDE$ in A and B, and by their mirror image mates $EDCBA$ and $EDCAB$ which also constitute a permuting set $EDC((AB))$ in A and B. Hence the order of projections on the 2-locus $H(A, B)$ is $(AB)CDE$. Similar inspections of open regions permuting in A and E show $B(AE)DC$ for $H(A, E)$, $BA(CD)E$ for $H(C, D)$, and so on.

Each of the r-loci axes will then have the property of zero distance among projections for at least r stimuli. The choice among these axes can be made by finding that axis which tends to parallel the major axis of the least intersection.

Orthogonality of Axes

A third desideratum in choosing a solution from among the exhaustive possibilities is that the axes chosen be more or less orthogonal to each other. Obviously, the r-loci will not in general be orthogonal, so that in a choice among these for the set of axes, only approximate orthogonality can eventuate at best. However, there is an advantage in choosing those r-loci which will have even this approximate degree of orthogonality, in that we may at least be sure that our description of the space is as nonredundant as possible.

Just as there is a feature of the space which allows one to approach the quasi-simple-structure criterion easily, so are there guides to orthogonality as well. If any set of r points in general position defines an r-locus line, the subspace of $r - 1$ dimensions in which the r points are embedded is everywhere orthogonal to the line of the r-locus. Also, if two points define a line in the space, then the 2-locus defines a hyperplane of $r - 1$ dimensions such that any line in the hyperplane is orthogonal to the line between the two points. For example, in Figures 1 and 2, note how each 2-locus H is a perpendicular bisector of, and hence orthogonal to, a line.

Now suppose that the first axis is chosen to be that r-locus which is approximately parallel to the major axis of the least intersection. The remaining $r - 1$ axes should be approximately parallel to the minor axes of the least intersection if they are to be orthogonal to the first; in other words, the remaining axes should be guided by the 2-locus which is orthogonal to the axis of the least intersection. In the example, A and E were, of course, found to define the major axis; thus, the 2-locus $H(A, E)$ is orthogonal to this line. By taking the second axis parallel to this 2-locus, one insures that it will be approximately orthogonal to the first axis: in this instance, the second axis must show A and E projecting on the same point. Actually, in the example, the only possible 2-locus which would then qualify as an axis would be $H(A, E)$ with order of projections $B(AE)DC$. With dimensionality higher than two, however, there would be a choice among a number of r-loci showing A and E adjacent in order, where A and E are the endpoints of the major axis.

The procedure for locating axes beyond the first follows the same general plan. Only r-loci showing the endpoints of the first axis adjacent are considered. Then, always omitting one of the stimulus points which served as endpoints for the first axis, that pair of stimulus points is found which serve as axis points for a least intersection enclosing the largest number of points which do not fall between the points in question on the first axis. For instance, in the example, if the four points A, B, C, D are considered, then A and D form the major axis for those four points. However, A and D are not taken as endpoints for the second dimension, since both B and C fall between A and D on the first axis. On the other hand, if B, C, D, E are examined, it is found that only the circle generated by B, C, and E has all four stimuli either conjoint or interior to the circle. In such a case, any of the pairs BC, BE, or CE could serve as the major axis for this subset of points. However, D falls both between B and E and C and E on the first axis, so that only B and C apparently qualify as endpoints on the second axis. This is a trivial finding for a two-dimensional example, of course, since there is only one rank order which qualifies in the first place. Nevertheless, the procedure would be the same for higher dimensionalities, always locating higher axes in terms of central intersections for reduced numbers of points such that there is minimal duplication of previous axes' rank orders.

The end results of an analysis based on these criteria would be a set of rank orders, representing projections upon axes which parallel major axes of least intersections, which show the property of quasi-simple structure, and which are approximately orthogonal to each other. Obviously it is not possible to plot such axes and perform any sort of rotational operations upon such a solution. All possible solutions are immediately at hand in such data, and if the solution obtained is unsatisfactory for some reason, there are certainly others which may be chosen by abandoning one or all of these criteria. However, the criteria proposed here do seem to have some recommendation on common-sense grounds as well as by analogy to current practice in factor analysis. On the other hand, only properties of the isotonic space itself are relied on in these criteria for selecting among the possible solutions, and these criteria are presented here as isotonic principles *sui generis*, and not as approximations to results which might be found by metric methods. The analogies drawn to principal axes and simple structure are meant only to be expository and suggestive rather than exact.

The Problem of Incomplete Data

Any method for multidimensional unfolding has very limited practical utility as long as it is limited to the case of complete data. The number of isotonic regions which may exist even for a small number of stimuli in small dimensionality grows truly astronomical very quickly with increase in either. As was pointed out in [2], complete data cannot possibly be obtained except

in the most trivial cases. Our purpose in limiting this discussion to the case
of complete data was, however, simply to make an exceedingly complicated
topic a jot more comprehensible. In applications to real data some com-
plexities do arise, mainly due to the fact that one does not necessarily have
the exhaustive solution already implicit in the data in the incomplete case,
and thus other steps must be introduced to supply the information given
in the complete case by the mirror image pairs. However, the general ideas
both for dimensionality and for a configurational solution may be applied to
the case of either complete or very incomplete data, so that the principles
enunciated here will be the basis for future discussion of the incomplete
data situation.

REFERENCES

[1] Bennett, J. F. Determination of the number of independent parameters of a score
    matrix from the examination of rank orders. *Psychometrika*, 1956, **21**, 383–393.
[2] Bennett, J. F. and Hays, W. L. Multidimensional unfolding: determining the dimen-
    sionality of ranked preference data. *Psychometrika*, 1960, **25**, 27–43.
[3] Coombs, C. H. Psychological scaling without a unit of measurement. *Psychol. Rev.*,
    1950, **57**, 145–158.
[4] Coombs, C. H. A theory of psychological scaling. *Engng. Res. Bull.* No. 34, Ann Arbor,
    Mich.: Univ. Michigan Press, 1952.
[5] Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univ. Chicago Press, 1947.

EMPIRICAL COMPARISON OF ITEM PARAMETERS
BASED ON THE
LOGISTIC AND NORMAL FUNCTIONS

FRANK B. BAKER

UNIVERSITY OF MARYLAND

Maximum likelihood estimates of item parameters of a scholastic aptitude test were computed using the normal and logistic models. The goodness of fit of ogives specified by the pairs of item parameters to the observed data was determined for all items. While negligible differences in the limen values were found, differences in item discrimination indices indicated that interpretation of these indices requires separate frames of reference. The empirical results showed the logistic model to be a useful alternative to the normal model in item analysis.

Parameters of mental test items typically have been based on the normal ogive model. The normal model has also been used in investigations of quantal response bioassay [8]. Berkson [2, 3, 4], however, has advocated the use of the logistic function in bioassay problems involving the fitting of an ogive curve to the observed proportions of response given by those possessing successively higher levels of the criterion. Recently Maxwell [16] has suggested the use of the logistic function rather than the normal in obtaining maximum likelihood estimates of mental test item parameters. Though the logistic function is a useful alternative to the normal function in toxicological investigations, comparable evidence is lacking in psychometrics. Empirical comparison of maximum likelihood estimates of item parameters based upon the logistic and normal functions have not been published. The availability of modern computing machines has made such an empirical investigation possible [1]. To provide a background for the present investigation, item parameters are defined and maximum likelihood methods for estimating these parameters are presented below.

The mental test theory developed by Lawley [12] is based upon maximum likelihood estimation of the parameters μ and σ of the normal ogive fitted to the proportions of correct item response, observed at the several ability levels. The parameter μ is known as the limen value—the point on the ability scale at which the probability of correct response is one-half. The reciprocal of the parameter σ is called the precision or discriminating power of the item. Since Lawley's initial application of maximum likelihood to the estimation of item parameters [12] several authors [5, 6, 7, 14, 15, 17] have modified and extended Lawley's basic theory. Although considerable mental test theory

is based upon maximum likelihood estimation of item parameters, practical applications of the method are rare. This author was unable to locate a published study in which parameters for items of a mental test were actually computed by the method of maximum likelihood. The reluctance to apply empirically this theoretically excellent technique is probably due to the laborious computational tasks involved.

The basic process in the maximum likelihood estimation of item parameters is that of fitting a two-parameter cumulative distribution function to the observed data. Finney [8] has shown this process is identical to fitting a linear regression line to the corresponding linear transformations of the proportions of correct item response observed at the several ability levels. Thus, the parameters α, β, of location and scale respectively, are estimated rather than the parameters of the function. In quantal response bioassay the ratio $- \alpha/\beta$, denoted by X_{50}, is the median lethal dose. The definition of the median lethal dose, the dosage level at which the probability of succumbing is equal to one-half, is analogous to that of the limen value in psychophysical measurement. For purposes of the present investigation the limen value is denoted by X_{50} and the discriminating power of the item by β. An index of item discrimination is better understood if the index increases as the discrimination increases; the scale parameter β has this property. The maximum likelihood estimates of the item parameters X_{50} and β are given by x_{50} and b.

$$(1) \qquad\qquad\qquad x_{50} = -a/b,$$

where a and b are the maximum likelihood estimates of the linear regression parameters α and β. The maximum likelihood estimates of these parameters can be found by an iterative solution of the following system of equations or δa and δb [10].

$$(2) \qquad \begin{aligned} \delta a \sum n_i w_i + \delta b \sum n_i w_i x_i &= \sum n_i w_i y_i . \\ \delta a \sum n_i w_i x_i + \delta b \sum n_i w_i x_i^2 &= \sum n_i w_i x_i y_i . \end{aligned}$$

When the normal function is used

$$(3) \qquad\qquad\qquad w_i = \hat{Z}_i^2/\hat{P}_i \hat{Q}_i ,$$

the weighting coefficient, and

$$(4) \qquad\qquad\qquad y_i = (P_i - \hat{P}_i)/\hat{Z}_i ,$$

the working value, where P_i is the observed proportion of correct response at the ability level x_i,

$$(5) \qquad\qquad \hat{P}_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a+bx_i} \exp\left(-u^2/2\right) du,$$

$$(6) \qquad\qquad\qquad \hat{Q}_i = 1 - \hat{P}_i ,$$

(7)
$$\hat{Z}_i = \frac{1}{\sqrt{2\pi}} \exp\left[-(a + bx_i)^2/2\right].$$

While tables are available which provide w and y for selected values of $a + bx_i$ [9], the lack of sufficient detail usually necessitates interpolating in tables of the normal integral to obtain the values of $\hat{P}_i$, $\hat{Q}_i$, and $\hat{Z}_i$. The values of w and y must be computed at each point of the ability scale within a given iteration. Unless the number of ability levels used is very small and the initial estimates a and b are close to α and β, the lengthy computations make this application of maximum likelihood economically impractical in most research situations.

The computational procedures for maximum likelihood estimation of item parameters using the logistic function are similar to those for the normal function. The equations to be solved iteratively are identical to those given above. Only the weighting coefficient w and the working value y are different. When the logistic function is used the following expressions result:

(8)
$$w_i = \hat{P}_i \hat{Q}_i,$$

the weighting coefficient,

(9)
$$y_i = (P_i - \hat{P}_i)/\hat{P}_i \hat{Q}_i,$$

the working value,

(10)
$$\hat{P}_i = \frac{1}{1 + \exp\left[-(a' + b'x_i)\right]}.$$

The prime notation is used to distinguish the estimates of the parameters of location and scale from those of the normal function.

A major computational advantage of the logistic function is immediately apparent. The cumulative area $\hat{P}_i$ is an explicit function of $a' + b'x_i$, whereas with the normal function $\hat{P}_i$ can be obtained only by interpolating tables or numerically integrating the normal function. The former is done when hand calculations are performed and the latter when digital computers are used. It is shown below that the explicit nature of the integral of the logistic function is an important factor when the cost of analysis is considered.

Though not mentioned by Maxwell [16] the goodness of fit of the ogive, specified by estimates of the item parameters, can easily be determined. The value of χ^2 can be obtained at any iteration in the following manner.

Normal function

(11)
$$\chi^2 = \sum_{i=1}^{k} \frac{n_i(P_i - \hat{P}_i)^2}{\hat{P}_i \hat{Q}_i},$$

but

$$w_i = \hat{Z}_i^2/\hat{P}_i\hat{Q}_i \; ,$$
$$y_i = (P_i - \hat{P}_i)/\hat{Z}_i \; ;$$

then

$$\chi^2 = \sum_{i=1}^{k} n_i w_i y_i^2$$

with $k - 2$ degrees of freedom, where k is the number of ability levels used.

Logistic function

(12)
$$\chi^2 = \sum_{i=1}^{k} \frac{n_i(P_i - \hat{P}_i)^2}{\hat{P}_i\hat{Q}_i} \; ,$$

but

$$w_i = \hat{P}_i\hat{Q}_i \; ,$$
$$y_i = (P_i - \hat{P}_i)/\hat{P}_i\hat{Q}_i \; ;$$

then

$$\chi^2 = \sum_{i=1}^{k} n_i w_i y_i^2$$

with $k - 2$ degrees of freedom.

The time-consuming computational tasks have been a barrier to the maximum likelihood estimation of item parameters by users of item statistics. The introduction of modern, high-speed digital computers has removed this barrier. Baker [1] has written a Univac "Scientific" computer program for maximum likelihood estimation of the item parameters X_{50} and β. This computer program has a capacity of (up to) 192 items and samples of (up to) 767 subjects. The availability of inexpensive maximum likelihood estimates of item parameters was instrumental in making the present investigation possible.

The data used in the current investigation consisted of a normally distributed sample ($\bar{X} = 40$, $S_x = 17$) of 499 cases, which had been scored on the first 72 items of the Minnesota Scholastic Aptitude Test, MSAT, [13]. The theory underlying maximum likelihood estimation of item parameters assumes the independent variate is a measure of the single underlying ability [12]. Common, though not necessarily correct, practice has been to substitute the total test score for this measure. This practice was followed in the present investigation. The procedure followed was to fit the normal and logistic function to the same set of observed data. Thus, the empirical utility of the normal and logistic function could be compared directly. Maximum likelihood estimates of α, β, X_{50}, and the value of chi square were calculated for each of the 72 items using both the logistic and normal functions. The analysis based on the normal function was performed using the available computer

program [1]. This computer program was then modified to employ the logistic function and the analysis repeated. The complete results of the analyses may be obtained from the author upon request.

Although Maxwell [16] considered the item parameters X_{50} and β based upon the logistic function to be similar to the normal function, the empirical results indicated the discriminating power of an item based on the logistic function has a numerical value differing from that based on the normal function. The discriminating power of an item is a function of the variance of the ogive fitted to the data. In the case of the normal function, the square of the parameter of scale is the reciprocal of the variance, $\beta^2 = 1/\sigma^2$. The same does not hold for the logistic function, $\beta^2 = \pi^2/3\sigma^2$. Thus the maximum likelihood estimate of the scale parameter of the normal function and the maximum likelihood estimate of the scale parameter of the logistic function are not estimates of the same quantity. Perhaps a different symbol, say λ, for the discriminating power of the item should be used when the logistic function has been employed.

Although the maximum likelihood estimates of the scale parameters lead to different results, the estimates of the limen value, $x_{50} = -a/b$, are the same for both functions. Negligible differences were observed for 45 of the 72 items. The few large differences noted were found in several items having large values of x_{50} (2.5–4.0). The observed differences would not appreciably affect the interpretation of the items. The empirical results suggest there is little difference in the values of x_{50} based on the logistic and normal functions. The close agreement of the observed value of x_{50} is due to the logistic and normal functions both being symmetric about their mid-points, which occurs when the exponential term $\alpha + \beta x$ is equal to zero. If the only consideration is the estimation of a point on the ability scale at which the probability of response is one-half, many symmetric functions could be employed. Such is the case in quantal response bioassay, where the primary objective is estimation of the median lethal dose. In item analysis, however, one must be concerned not only with the value of x_{50} but also with the discriminating power of the item.

A pertinent question is, "Which of the two functions is the better model for representing the observed data?" The goodness of fit of the ogive curve, specified by x_{50} and b, to the observed item data can be tested by means of the chi-square criterion. The values of chi square, based on fitting both functions, were computed and are presented in Table 1. The ogive curves, specified by the pairs of item parameter estimates, differed significantly from the observed data in only 8 of the 72 items. One item (number 65) differed at the .05 level for the normal function and was nonsignificant for the logistic function. One item (number 15) was significantly different at the .01 level for the normal curve and at the .05 level for the logistic curve. The remaining 6 discrepant items had values of chi square greater than the .01 significance

TABLE 1

Chi Square Goodness of Fit Tests Based on the Normal and
Logistic Function for 72 Items of the Minnesota
Scholastic Aptitude Test

Item Number	Normal x^2	Logistic x^2	Item Number	Normal x^2	Logistic x^2
1	41.92	41.41	37	38.77	37.82
2	89.13**	88.31**	38	58.15	58.00
3	71.55	71.24	39	57.17	57.54
4	64.05	64.02	40	53.20	52.76
5	48.67	48.77	41	38.56	36.89
6	57.33	57.11	42	62.45	60.21
7	58.49	58.16	43	51.81	50.75
8	62.41	60.02	44	53.53	52.02
9	72.46	71.00	45	72.18	68.61
10	63.81	63.34	46	63.00	61.71
11	57.15	53.06	47	65.22	63.32
12	51.95	55.38	48	48.96	49.06
13	90.23**	85.37**	49	66.09	62.70
14	53.33	51.23	50	64.52	64.70
15	90.18**	81.81*	51	67.06	67.17
16	53.79	53.28	52	74.14	71.68
17	59.12	54.36	53	90.30**	89.14**
18	65.24	63.17	54	46.68	46.55
19	54.53	54.34	55	53.71	53.68
20	49.93	50.13	56	45.39	44.94
21	64.61	64.37	57	63.97	63.11
22	48.60	48.45	58	55.71	53.33
23	54.70	54.64	59	74.09	71.24
24	103.89**	98.53**	60	55.19	53.87
25	73.43	67.04	61	58.51	56.29
26	47.59	46.92	62	40.63	40.77
27	69.29	68.28	63	79.32	63.17
28	50.66	49.13	64	67.55	54.94
29	55.39	55.15	65	83.61	80.32
30	77.21	75.74	66	53.39	52.08
31	63.62	63.20	67	50.40	49.95
32	68.77	67.48	68	59.12	58.30
33	96.55**	91.61**	69	58.71	57.97
34	54.36	55.27	70	63.65	60.39
35	42.92	42.92	71	42.90	43.63
36	61.17	60.04	72	127.75	96.23**

**Significant at the .01 level for 62 degrees of freedom
*Significant at the .05 level for 62 degrees of freedom

level for both functions. From an empirical point of view there is little difference in the manner in which the logistic and normal models represent the item data used in the present study.

The logistic function possesses a distinct economic advantage over the normal function. In order to maintain sufficient accuracy in the solution of the normal equations, the values of w and y accurate to six decimal places were necessary. The values of $\hat{P}_i$, $\hat{Q}_i$, $\hat{Z}_i$ based on the normal function were obtained by numerical integration of that function. A normal table of comparable accuracy would have greatly exceeded the capacity of the computer. The values of $\hat{P}_i$ and $\hat{Q}_i$, based on the logistic function, were obtained directly from (10) using a series expansion to evaluate e^z. The computation of the 72 maximum likelihood estimates of X_{50} and β, based on the normal model, required one and one-half hours of computer running time. When the logistic function was employed, the computer running time was only one-half hour. With computer operating costs ranging from 100 to 500 dollars per hour, the saving of one hour of running time per analysis is an important budgetary consideration.

Conclusions

1. The use of the logistic model as an alternative to the normal model in maximum likelihood estimation of mental test item parameters had been suggested by Maxwell [16]. Such substitution appears empirically possible if separate frames of reference regarding the interpretation of the numerical values of the discrimination index β, parameter of scale, are used.

2. Due to the symmetry of the normal and logistic functions about their midpoints, very similar values of x_{50} were obtained.

3. The goodness of fit tests suggested little advantage for either the normal or logistic ogive fitted to the observed item data.

4. The computer running time showed the logistic function has a distinct economic advantage over the normal function. The cost of analysis using the former is approximately one-third that of the latter.

Though the logistic function can be used as an alternative to the normal function in empirical investigations of test items, its role in mental test theory needs to be investigated. To be useful in mental test theory, it must be demonstrated that other mental test constructs can also be derived mathematically using the logistic model. The employment of the logistic function in further extensions of Lawley's [12] work should prove fruitful. Birnbaum [5, 6, 7] has recently made several important contributions in this regard.

REFERENCES

[1] Baker, F. B. Univac Scientific computer program for test scoring and item analysis. Washington, D. C.: Amer. Documentation Inst., Library of Congress, Document 5931.

[2] Berkson, J. Why I prefer logits to probits. *Biometrics*, 1951, **7**, 327–339.

[3] Berkson, J. A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based upon the logistic function. *J. Amer. statist. Ass.*, 1953, **48**, 565–599.

[4] Berkson, J. Tables for the maximum likelihood estimate of the logistic function. *Biometrics*, 1957, **13**, 28–34.

[5] Birnbaum, A. Probability and statistics in item analysis and classification problems. Efficient design and use of tests of mental ability for various decision-making problems. Ser. Rep. No. 58–16, USAF Sch. Aviation Medicine, Randolph AFB, Texas, 1957.

[6] Birnbaum, A. Probability and statistics in item analysis and classification problems. Further considerations of efficiency in tests of a mental ability. Ser. Rep. No. 58–17, USAF Sch. Aviation Medicine, Randolph AFB, Texas, 1957.

[7] Birnbaum, A. Probability and statistics in item analysis and classification problems. On the estimation of mental ability. Ser. Rep. No. 15, USAF Sch. Aviation Medicine, Randolph AFB, Texas, 1957.

[8] Finney, D. J. The application of probit analysis to the results of mental tests. *Psychometrika*, 1944, **9**, 31–39.

[9] Finney, D. J. and Stevens, W. L. A table for calculating of working probits and weights in probit analysis. *Biometrika*, 1948, **35**, 191–201.

[10] Garwood, F. The application of maximum likelihood to dosage-mortality curves. *Biometrika*, 1941, **32**, 46–58.

[11] Hodges, J. L., Jr. Fitting the logistic by maximum likelihood. *Biometrics*, 1958, **14**, 453–461.

[12] Lawley, D. N. On problems connected with item selection and test construction. *Proc. Roy. Soc., Edinburgh*, 1943, **61**A, 273–287.

[13] Layton, W. L. Construction of a short form of the Ohio State University psychological examination. Univ. Minnesota, Student Counseling Bureau, 1957.

[14] Lord, F. A theory of test scores. *Psychometric Monogr. No. 7*, Chicago: Univ. Chicago Press, 1952.

[15] Maritz, J. S. Estimation of the correlation coefficient in the case of a bivariate normal distribution when one of the variables is dichotomized. *Psychometrika*, 1953, **18**, 97–110.

[16] Maxwell, A. E. Maximum likelihood estimates of item parameters using the logistic function. *Psychometrika*, 1959, **24**, 221–227.

[17] Tucker, L. R. Maximum validity of a test with equivalent items. *Psychometrika*, 1946, **11**, 1–14.

BOOK REVIEWS

Don Lewis. *Quantitative Methods in Psychology.* New York: McGraw-Hill Book Co., Inc., 1960. Pp. xii + 558.

"Graduate students of psychology, with or without college training in mathematics, ... usually have not formed habits of thinking clearly and consistently in quantitative (functional) terms, and they do not acquire such habits until they have been tutored extensively in the use of mathematics as a tool of scientific inquiry." In accordance with this belief, Professor Lewis developed a graduate course in quantitative methods some 20 years ago. The present text for the course started as a supplement for Daniels' *Mathematical Preparation for Physical Chemistry*, and grew into a paper-bound book of about 300 pages by 1948. The present edition has added more than 200 pages.

Chapter 1 starts with definitions of *variables, functions,* and other basic terms. Various methods for fitting straight lines are developed in Chapter 2. Logarithms and their uses are defined and described in Chapter 3. Chapter 4 deals with fitting nonlinear functions. From here on, the text is at a distinctly greater level of difficulty. The basic rules for differentiation and integration are covered in the next two chapters. Chapters 7, 8, and 9 apply these rules in deriving and describing several fundamental statistical distributions—the normal curve, the Poisson, chi square, t, and F. Chapter 10 gives examples of appropriate and inappropriate use of chi square as a goodness of fit statistic and goes on to suggest F-ratio tests of linearity for independent and dependent observations, and the use of orthogonal polynomials for fitting nonlinear trends. Each chapter is followed by exercises for the student; good problems are provided for every topic.

Chapter 11, the last chapter, is the most satisfying. It is a lengthy, meaty, historical discussion of applications of mathematical thinking by such psychologists as Thurstone, Hull, Stevens, Estes, and Burke. A refreshing freedom of style and thought lightens this critical account of attempts to find universal laws of behavior.

Another section that students will find both novel and useful is Chapter 4 on curve fitting. Many texts show how logarithms and other elementary transformations will change parabolas, hyperbolas, exponentials, etc., into straight lines when the intercept is zero. Lewis deals with the more difficut approximations that are necessary when these nonlinear equations have an additive constant. Useful hints and principles are scattered through the chapter. He is careful to make the following points: minimizing the deviance of log Y from its predicted value is not the same as a least-square regression of Y on X; least-square equations can be derived for nonlinear functions; approximate solutions for nonlinear equations can be found by using guessed values for all unknowns but one, and iterating.

Because his emphasis is on quantitative methods, Lewis does not refer to the abstract algebra of sets, lattices, groups, etc., which has become so important in the exact sciences. Those who believe that abstract algebra might also be useful for the inexact sciences will regret the omission.

Some 215 pages of this 558 page book are devoted to statistics, 130 pages to the derivation of standard statistical distributions, and about 85 pages to goodness of fit and analysis of variance for trends.

Most of the 130 pages could be justified on the basis that detailed derivations of the basic sampling distributions are not easily available, and this text will serve as convenient reference. But was it really necessary to illustrate in drawn-out detail the application

of the chi square to frequencies, the application of the t-test to correlated and uncorrelated measures, as well as the analysis of variance for one-way and two-way classification?

Chapter 10, on goodness of fit and trend analysis, does not appear to contribute to the aim of the book. At best, it is a duplication of standard innocuous material. But, unfortunately, Lewis has succeeded in including two of the standard errors (so to speak) found in Lindquist, McNemar, Edwards, etc. (a) The fallacious notion that using "subjects" as a classification factor in analysis of variance eliminates the effects of dependence and correlation between observations. (b) The misleading idea that F-ratio tests are inapplicable whenever Bartlett's test shows that the cell variances are unequal; Edwards has called attention to this mistake in his 1960 revision of *Experimental Design*.

In general, the strategic and tactical principles advocated by Lewis are admirably right and reasonable. "Any estimate of a . . . constant regardless of how it is obtained, should make sense" (p. 59). "In a sense [curvefitting] is an art . . . There is almost always room for disagreement and for judicious decisions" (p. 115). "We would ordinarily not be interested (or even justified) in making any tests concerning any [polynomial term] for which we lack a specific rationale" (pp. 415-416). But occasionally there is a slip in applying these principles.

Chapter 10 is marred by two such examples of weak statistical tactics. The first muddle occurs when Lewis attempts to compare the trends for four groups. Each group has 20 subjects and each subject has 5 trials. Having found that the group-by-trial inter-action term is significant by the F-ratio test and fitted a slope coefficient to each group trend, he then becomes disturbed about the fact that within each group, the 5 trial scores are interrelated. "This means that we have no satisfactory way of estimating the sampling variabilities of the regression coefficients . . ." (p. 397). Since Lewis treats each group profile as a straight line, there is a fairly clear way of obtaining the sampling variance of each group slope, b_g. If we fit a slope, b_i for the ith S in group g, then $\bar{b}$, the average slope for all subjects in group g, is equal to b_g. It follows that within group g the variance of the b_i's about $\bar{b}$ is the variance of b_g. In fact, a one-way analysis of variance, substituting b_i for each vector of 5 trials scores can be carried out just as if b_i were a regular score. Instead of this Lewis carried out 6 separate trend analyses.

Lewis has succumbed here to the common, but false, belief that the group-by-trial interaction term is an appropriate test for the parallelism of group profiles. Greenhouse and Geisser (*Psychometrika*, 1959, **24**, 95-112) have extended the 1954 paper of G.E.P. Box on one group, to the case of several groups and shown that generally, use of the interaction term inflates the significance of any difference. (Incidentally, they state "As is well known, in order that the usually computed ratios of mean squares in this model be exactly distributed as the F distribution, it is necessary that columns, in addition to being normally distributed, have equal variances and be mutually independent or, at most, have equal correlations." Presumably they are being ironic. I can't find any statistical psychology text that displays knowledge of this assumption and a glance at the pages of the *Journal of Experimental Psychology*, or any other respectable psychological journal, will show an equally blissful ignorance of this assumption.)

If the question can be answered in terms of a single score for each subject, derived from the set of raw observations for each subject, then independence of derived scores is guaranteed if the subjects have been sampled independently. Reducing the vector of scores for each subject to one derived score also has the convenience of eliminating the group-by-trial interaction term.

Another example of weak tactics is given in a demonstration of the use of orthogonal polynomials (pp. 409-417). The data were taken from a generalization experiment in which Grant and Schiller conditioned GSR responses to a lighted rectangle 1 inch wide and 12 inches high. Subjects were tested for generalization with rectangle heights 9, 10, 11, 12, 13, 14, and 15 inches. Grant and Schiller expected that GSR response would decrease

as the rectangle height deviated from 12 inches. Experimental steps were taken to eliminate tendencies for the higher rectangles to cause an increased response (to the larger areas).

The specific a priori equation that can be deduced from these considerations is $Y = w_0 + w_1(X - 12)^2$, where w_0 and w_1 are arbitrary constants, Y is the GSR response, and X is the height of the rectangle. The proportion of deviance due to the a priori quadratic equation given above is .052. The squared correlation ratio due to height is .128. An F-ratio test shows there is no significant difference between the proportion of deviance accounted for by using the best least-square curvilinear fit to the average GSR response at each height, and the a priori quadratic function. The expectations of Grant and Schiller are satisfied very nicely by their data.

Instead of first testing the quadratic function, Lewis fits each orthogonal polynomial in turn. After fitting the linear term he says "this value [of the linear term] is associated with a probability of about .09. We find ourselves in the somewhat strange predicament . . . of neither accepting nor rejecting linear regression. We shall, nevertheless, proceed with the analysis. . ." (p. 415).

Statistical analysis is often woefully uncertain. It isn't often that we have a straightforward equation deduced from a priori considerations. Lewis implies several times in his text that rational equations should be applied first and then the residual variance examined to see if anything signficant is left. If he had followed this principle, it would have become clear immediately that the quadratic equation accounts, within statistical limits, for all significant variation. This particular bit of harassment for the student is really unnecessary for, on p. 417, Lewis states firmly that "only the quadratic component contributes significantly to regression."

Various minor confusions in notation and concept have found secure niches in the statistical portions of Lewis' text. For example, the method of least squares is recommended for "maximum precision" under all circumstances (p. 19) without explaining that it may lose a great deal of that precision when the deviations have a non-normal distributions Again on p. 180, Lewis develops a maximum likelihood solution for the regression coefficientd in a bivariate normal distribution. It is emphasized that the goal is to obtain unbiaset estimates of the regression coefficients. Maximum likelihood estimates, however, are no. necessarily unbiased estimates unless certain restrictions are met.

The notation for variance is inconsistent. The symbol σ^2 is used on p. 26 and elsewhere to represent the biased sample estimate, $\sum_{i=1}^{N}(X_i - \bar{X})^2/N$. On p. 291, σ^2 is the true population variance and s^2 is used for the biased sample estimate. On p. 300, s'^2 is used to denote the usual unbiased sample estimate $\sum_{i=1}^{N}(X_i - \bar{X})^2/(N-1)$, but on p. 363, σ^2 is used for the unbiased sample estimate of a population variance! Usually this slippery notation is used dextrously enough but Lewis comes a cropper at least once when he gives an underestimate of the population error variance about a line (p. 31). Another minor annoyance is the consistent lack of subscripts with the summation signs. The reader is rarely told whether the sum is to be taken over the individual, the cell, all the cells, or whatever.

Clearly the book is worth buying for teachers and practitioners of experimental psychology (with the possible exception of the Skinner cult) if only for Chapter 4 on curvefitting and Chapter 11 on attempts to find quantitative laws in psychology. But for class use, it will take an ever-vigilant instructor who can pilot his students around the pitfalls, point out statistical fallacies when they occur, and substitute direct attacks for some tangential approaches.

I hope that the next revision will decrease the large amount of space devoted to statistical techniques and examples that are alien to the theme of the book.

Walter Reed Army Institute
  of Research
Washington, D. C.

ARDIE LUBIN

W. S. Ray. *An Introduction to Experimental Design.* New York: The Macmillan Company, 1960. Pp. 254.

It is refreshing to meet a writer of an introductory textbook on psychological statistics who shows more concern with the logic of inductive inference and its close links with the design of experiments than with the arithmetic of statistics. When, in addition, he refers to original sources for the basic statistical theory this too is greatly to his credit. Dr. Ray is such a writer.

In his *Introduction to Experimental Design* his declared aim is "to start with the simplest ideas and principles of design and analysis, to take up in proper order a few central developments...and to end with certain interesting and advanced topics and issues." The first part of this aim is realized in nine chapters covering 108 pages of text. Although classical terminology, perhaps unwisely, is avoided, this part deals mainly with one- and two-way analysis of variance and the familiar randomized blocks type of design. Latin squares are mentioned, but only in passing. The emphasis throughout is on the correct planning of experiments to enable unbiased comparison of treatment effects to be made. Methods of increasing precision by matching and by the elimination of known components of variability—and later by adjusting—are brought to the reader's attention in a vivid way. Yet this early section of the book is a bit prolix and repetitive. Also, some basic notions, such as degrees of freedom and precision, are introduced without adequate discussion, while the author's persistent reference to individual differences—which are very real effects—as error rather than within group variation is unrealistic. Noticeably missing, too, are references to the sequential nature of much psychological data, while on page 104 *et seq.* a fine opportunity is lost to mention the value of simple cross-over designs for dealing with order effects.

Chapter 10, entitled "Adjusting," introduces the reader to linear regression and analysis of covariance; the latter topic is expanded in chapter 13. The intervening chapters (chapters 11 and 12) give a good introduction to factorial designs. Here the use of classical terminology is probably very wise; although perhaps a bit forbidding to the beginner, it will prepare him for further reading. Chapters 14 through 16 elaborate topics which naturally arise from factorial arrangements, and the book ends with a chapter on miscellaneous points of interest including a mention of how to deal with situations in which missing readings occur.

All things considered it seems fair to say that Dr. Ray's book is a useful addition to the short list of textbooks on experimental design suitable for beginners and for private reading. The earlier part of the book is a little verbose, while after chapter 10 the difficulty increases perhaps too abruptly. In the latter section insufficient attention is paid to a realistic interpretation of the results of the analyses performed, but in this respect the author was gravely handicapped by his liberal use of artificial data.

Maudsley Hospital   A. E. Maxwell
London, England

Communications Biophysics Group of Research Laboratory of Electronics and William M. Siebert. *Processing Neuroelectric Data.* Cambridge: Massachusetts Institute of Technology, Research Laboratory of Electronics, Technical Report 351, 1959. Pp. vii + 121.

Many people concerned with the recording and analysis of neuroelectric data are statistically naive. This monograph will help to remedy this state of affairs. It consists of four chapters and four appendices. Anyone unfamiliar with statistical methods should

begin by reading the first two appendices in which random processes are described and mathematical statistics introduced. These two appendices provide a valuable introduction to statistical methods applicable to neuroelectric data. In the remaining two appendices computers are described and a selected list of the group's publications given.

The first (introductory) chapter deals with the quantification of neuroelectric data in general. It is stated that the group responsible for the monograph have developed certain methods of data processing and certain types of mathematical models which are "capable of coming to grips with the statistical character of neural activity which is one of the essential features of the nervous system." It is these techniques which are described in the monograph.

The second chapter, "Evoked Responses," presents some quantitative descriptions of evoked responses as recorded by gross electrodes from many cells. This chapter is mostly concerned with the averaging of such responses by analog and digital computers. But the measurement of other statistics, such as dispersion and skewness, is also briefly considered.

The third chapter deals with two techniques for the processing of EEG data, i.e., spontaneous potentials recorded from the scalp. One technique represents an effort to perform electronically a type of analysis that is similar to that carried out by an electroencephalographer when he visually examines the EEG for certain rhythmic characteristics, such as the occurrence and number of rhythmic bursts. The other technique uses correlation analysis in the study of the EEG.

The monograph concludes by stressing the need to bring about a rapprochement between biological studies of the nervous system and studies in which neural behavior is simulated on computers. Such a rapprochement, it is hoped, may lead to the provision of catalogues of possible mathematical models. One of the aims of the monograph is to convince young researchers from the physical and life sciences of the desirability of acquiring the necessary skills. If the monograph is read with sufficient care, this aim should be achieved.

The authors are to be congratulated on producing a clear and concise account of the processing methods they have successfully used.

R. F. GARSIDE

University of Durham

O. HOBART MOWRER. *Learning Theory and Behavior*. New York: John Wiley and Sons, Inc., 1960. Pp. xiv + 555.

This book, in conjunction with a companion volume *Learning Theory and the Symbolic Processes*, presents Professor Mowrer's latest formulations. This theoretical edifice has been built on a foundation of over 20 years of investigation of the learning process, much of it pioneering in nature. It is therefore not surprising that Mowrer's presentation is both scholarly and provocative. Following an abbreviated history of learning theory, in which cognitive theorists are noticeably absent, Mowrer goes on to describe the shortcomings and inadequacies of the early efforts, the revisions in learning theory precipitated by these inadequacies, the now classical disputes which arose over these revisions, and the remaining shortcomings of modern learning theory.

This historic development culminates in the author's presentation of his revised two-factor theory, in which all learning is reduced to classical conditioning or sign learning. The two factors now refer to the forms of reinforcement involved, incremental and decremental, rather than to two types of learning. Central to this position is the view that a number of intervening states, e.g., the emotions of fear, disappointment, relief, and hope, become conditioned to particular independent or response-correlated stimuli and it is the increment and decrement of these states which reinforce particular behavioral acts. Since

such internal states are conceptualized as secondary drives or primary drive derivatives (e.g., hunger-fear) and since these states are viewed as mediating all learning, primary attention is paid to secondary reinforcement. Mowrer's analysis of secondary reinforcement leads him to the intriguing conclusion that it is synonymous with the concept of habit. Thus, to Mowrer, habit or the conventional S-R bond is reducible to those independent or response-initiated stimuli which, through conditioning, instigate particular emotional states, e.g., fear or hope, which in turn stimulate the organism to particular responses.

Professor Mowrer's argument, while always stimulating, does present certain diffi-culties. Not the least of these is the scant evidence presented that shock offset could be used to condition hope or that fear could be conditioned by means of the induction of metabolic drives. Another matter of concern is Mowrer's rejection of the fractional goal reaction in favor of such concepts as hope. The reviewer applauds the inclusion of such concepts within a theoretical system and feels that perhaps Mowrer's greatest contribution is his conception of the emotions as psychologically meaningful states, central in importance, rather than physiological entities, disruptive events, or mere epiphenomena. However, the reviewer must confess to some confusion concerning Mowrer's usage of such concepts. Mowrer explicitly states that hope and the other emotions are to be identified in terms of the situations that presumably evoke them rather than in terms of subjective descriptions or physiological concomitants. Nevertheless he frequently lapses into discussions which suggest that the emotions are to be identified with both the states of the autonomic nervous system and with cognitive events. This latter identification becomes especially apparent in Mowrer's equation of hope and fear with Tolman's concept of expectations.

Other exceptions will be taken to this book. Certain readers will object to Mowrer's somewhat cavalier handling of particular thinkers and issues, his treatment of homeostasis as a concept possessing specific predictive ability rather than being assumptive in nature, his sometimes superfluous neurologizing, and above all, certain stylistic shortcomings in Mowrer's presentation which make for very difficult reading. These criticisms notwith-standing, there is little doubt that this book represents a major contribution to our under-standing of both the learning process and behavior in general.

Yale University   EDWARD ZIGLER

MARSHALL B. JONES. *Simplex Theory*. U. S. Naval School of Aviation Medicine, Monograph Series No. 3. Pensacola, Florida: U. S. Naval Aviation Medicine Center, June 1, 1959. Pp. v + 106.

This monograph consists of four chapters, each with an appendix of supplementary notes designed to elaborate further or prove some of the assertions in the text. The subject with which the monograph purportedly deals is simplex theory. By simplex the author means "... a sequence of stages each one of which is stacked within the next like the sections of a telescope" (p. 11). This definition indeed sets the pattern for the entire mono-graph—it is for the most part a collection of sweeping generalizations, false analogies, and circular arguments. This remark is directed at the "theory" aspect of the study only. It is neither a denial nor a minimization of the importance of the concrete problem under investigation—the design and testing of the best sequence of courses for naval air training which involves many complex tasks such as flying combat planes during daytime or at night, under instrument or visual fly-conditions, etc. However the particular conclusions reached for the best sequence of such a training program may well be drawn from actual experience with the problem at hand and not from the loose theory offered. Basically, the

author's molar correlational analysis is nothing more than Guttman's scalogram analysis, and his simplicial form is no different from Guttman's perfect scale.

Chapter 1 is entitled "Molar Correlational Analysis." What this is precisely is only illustrated but never really said. As far as this reviewer is able to determine, a molar correlational pattern exists among a set of correlations when (1) one can find more than one focus, and (2) the foci are independent (p. 5). This loose analogy is alleged to offer a superior alternative model to factor analysis of the same data.

Chapter 2, "Simplex and Simplicial Form," starts with a metaphoric definition and a rather confused discussion on "simplex" and "simplicial form," between which even the author himself admits to be ". . . at some pains to distinguish . . ." (p. 57). The difference is that "a simplex is a hypothesis, while a simplicial form is an observable fact" (p. 57). The remaining part of this chapter is devoted to a review of factor analysis under the caption: "And Some History," and the review is done clearly with a view to showing that factor analysis is inferior to molar correlational analysis. For example, in a learning experiment, the author feels ". . . that by no amount of factor analysis can we ever legitimately conclude that the later stages are simpler. This conclusion follows only from the molar pattern of the practice matrix. Once we recognize this pattern the underlying structure is easily reached—but only by hypothesis. If we insist upon computing our theories, we will never get there" (p. 22). The same point is emphasized later: "All these things depend upon the recognition of molar pattern, and for this we need to look at the matrix, not factor it" (p. 35).

Chapter 3 is devoted to what is presumed to be the author's major contribution: "The String Model." Actually, it consists of elementary considerations of graph and lattice theory, all couched in a peculiar jargon of the author's own invention. For example, a 0-cell in a directed graph is called a "root" if it is an "original" or leftmost point, a "terminal" point if it is a rightmost point, or a "passage" point if through it a "trunkline" passes. A "trunkline" is a connected chain of 1-simplices from an original to a terminal point. A bounding 0-cell is a singly connected point, that on a 1-cycle is a doubly connected point, etc. The "body" of a model ". . . always intervenes between its roots and its branches, which is, of course, precisely as it should be," (p. 48) and bodies are classified further into "hub," "stock," "frame," and "middle." The discussion is further marred by the author's failure to distinguish between a definition and a metaphor. For example, ". . . we must first recognize that the notion of a trunk is not a single idea. Nevertheless, I think, any reasonable conception of a trunk has something in common with any other; and that 'something in common' is what we have defined as the stock of the tree" (p. 50). Or, "essentially, a frame means that we don't really have a tree but a bush, perhaps, or a vine, or at any rate something in which roots do not necessarily precede branches" (p. 53). And, ". . . the middle is the only one of the four kinds of body which has a 'seam' " (p. 53).

The last chapter "How To Do It" supposedly offers a scheme for finding the solution (by no means unique, however) to any simplicial analysis, but the question as to whether or not this can be done ". . . without reference to the content of the stages" is answered with both yes and no (p. 59). If the stages do not satisfy a simplicial form, "reliability" or "partial commitment" is introduced to force them into such a structure (p. 59). All in all, the reader can hardly keep up with the constant appearance of new words, and if anything fails to fit the model, a new concept is injected at once to "explain" away the discrepancy. Apparently, this chain of argument continues without much restraint or objective, and it culminates with the somewhat disappointing remark that ". . . in sufficient detail, every simplicial relationship must fall. Now, while the proposition is incontestable, it is not disastrous. Our situation is, perhaps, best approached through an analogy . . ." (p. 70).

The appendices consist mostly of prolonged belaboring of either some trivialities or downright false assertions, and no clear distinction is ever made between a definition and a proof. For example, an "uncrossed" model is a model ". . . in which it is possible

to make every line straight without crossing any two lines except at a point which is in the model" (p. 93). Or, "... so long as the points of attachment straddle one another the model is hopelessly crossed" (p. 94). Finally, most proofs do not advance much beyond the level exemplified by the following: "... to prove that every reduced model is singly connected. If any two ordinary points were multiply connected, the model would not be reduced ..." (p. 98).

A word may be added concerning the style, which is largely in the form of conversational English. The organization is mostly discursive and often amorphous. This weakness could well be the result of a rushed job, as the author himself pointed out in the preface.

System Development Corporation   RICHARD C. KAO